QR1. I4

STIC-ILL

From: Goldberg, Jeanine
Sent: Friday, December 01, 2000 2:57 PM
To: STIC-ILL
Subject: please pull HCV

1. Proc. Natl. Acad. Sci. U.S.A. 91 (21), 10134-10138 (1994)
2. J. Virol. 66, 3225-3229 (1992)

3. Hepatology 19 (3), 551-553 (1994)

4. J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)

THANK YOU

Jeanine Enewold Goldberg
1655
CM1--12D11
306-5817

Journal of General Virology (1992), 73, 1131–1141. Printed in Great Britain

Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants

S.-W. Chan,[1] F. McOmish,[2] E. C. Holmes,[3] B. Dow,[4] J. F. Peutherer,[1] E. Follett,[4] P. L. Yap[2] and P. Simmonds[1]*

[1]Department of Medical Microbiology, Medical School, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, [2]Edinburgh and South East Scotland Blood Transfusion Service, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9HB, [3]Division of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JN and [4]SNBTS Microbiology Reference Unit, Regional Virus Laboratory, Ruchill Hospital, Glasgow G20 9NB, U.K.

Sequences obtained in the 5′ non-coding region (5′NCR) of hepatitis C virus (HCV) were obtained from Scottish blood donors and compared with previously published HCV sequences. Phylogenetic analysis revealed the existence of three distinct groups of sequences; two of these corresponded to the recently described HCV types 1 and 2 variants, while viral sequences detected in around a third of the blood donors formed a separate phylogenetic group that probably represents infection with a novel virus species. Nucleotide sequences of this latter group differed from all previously published 5′NCR sequence variants by at least 9%. This new virus type also differed considerably from previously published variants in other regions of the viral genome (core, NS-3 and NS-5), with corrected nucleotide distances of 15, 43 and 49% respectively from the prototype HCV-1 sequence. Formal phylogenetic analysis of each of the coding regions confirmed that HCV type 1 variants could be clearly differentiated into regional variants (Far East and U.S.A./European), in contrast to the clearly overlapping geographical distributions of the main HCV types in U.K. blood donors. We discuss the evidence for and against the hypothesis that the three main phylogenetic groups identified in this study represent separate species of HCV.

Introduction

The aetiological agent responsible for most cases of post-transfusion non-A, non-B (NANB) hepatitis has been cloned and characterized (Choo *et al.*, 1989, 1991). This virus, now termed hepatitis C (HCV), is a positive-strand RNA virus distantly related to the pestiviruses and flaviviruses (Miller & Purcell, 1990; Koonin, 1991). Its genome consists of an approximately 332 nucleotide 5′ non-coding region (5′NCR) followed by a continuous single open reading frame encoding a polypeptide of around 3010 amino acids and then a short 3′ untranslated region (Kato *et al.*, 1990; Takamizawa *et al.*, 1991; Choo *et al.*, 1991). By analogy with flaviviruses, this polypeptide has been divided into a 5′ structural region consisting of putative core and envelope proteins and a 3′ region corresponding to non-structural (NS-1 to NS-5) proteins. Recombinant proteins cloned from the prototype virus and synthetic peptides based on the viral sequence have been used to detect HCV antibodies (Kuo *et al.*, 1989; Muraiso *et al.*, 1990; Hosein *et al.*, 1991), and screening of blood donors has been initiated in several countries to prevent post-transfusional NANB hepatitis. However, there remain donations that transmit hepatitis C virus but which are seronegative or 'indeterminate' in commercial serological tests (Esteban *et al.*, 1990; van der Poel *et al.*, 1991; Japanese Red Cross Non-A, Non-B Research Group, 1991). It is possible that some of these false negative serological results may be the result of infection by extreme sequence variants of HCV that elicit an antibody response that has limited or no cross-reactivity with the peptide antigens used in serological assays. Supporting this hypothesis is the recent discovery of HCV variants (Enomoto *et al.*, 1990; Nakao *et al.*, 1991; Okamoto *et al.*, 1991) that differ markedly in sequence from the original prototype HCV (HCV-1; Choo *et al.*, 1991) and others found in Japanese patients (Kato *et al.*, 1990; Takamizawa *et al.*, 1991).

To investigate whether sequence heterogeneity might influence the effectiveness of serological screening for HCV in blood donors, we initiated a study to examine nucleotide sequence diversity of HCV in naturally

The nucleotide sequence data reported in this paper have been assigned GenBank accession numbers D10113 to D10134.

infected individuals. Available for the study were anti-HCV positive blood donors, intravenous drug users (IVDUs) and haemophiliacs exposed previously to non-heat-treated factor VIII, and who had biochemical evidence of liver disease (Simmonds *et al.*, 1990*b*). It is hoped that such studies will also assist in the development of type-specific and type-common antigens for serological diagnosis, to allow the detection and typing of HCV variants by polymerase chain reaction (PCR), and may ultimately assist in vaccine research.

Methods

Samples. Plasma from 18 different blood donors (E-b1 to E-b18), that were repeatedly reactive on screening by Abbott second generation enzyme immunoassay (EIA), and confirmed or indeterminate by a recombinant immunoblot assay (RIBA, Ortho; Chan *et al.*, 1991) were the principal samples used in this study. Sequences in the NS-3 region from five anti-HCV positive IVDUs (abbreviated as i1 to i5 in Simmonds *et al.*, 1990*b*), five haemophiliacs who had received non-heat-treated clotting concentrate, and who were also anti-HCV-positive (h1 to h5), three pools of 1000 donations collected in 1983 (p1 to p3), and five separate batches of commercially available non-heat-treated factor VIII (f1 to f5) correspond to those described previously (Simmonds *et al.*, 1990*b*).

Primers. The primers used for cDNA synthesis and PCR are listed in Table 1. They were synthesized by Oswel DNA Service, Department of Chemistry, University of Edinburgh, U.K.

RNA extraction and PCR. HCV virions in 0·2 to 1·0 ml volumes of plasma were pelleted from plasma by ultracentrifugation at 100000 *g* for 2 h at 4 °C. RNA was extracted from the pellet as previously described (Chomczynski & Sacchi, 1987; Simmonds *et al.*, 1990*b*). First strand cDNA was synthesized from 3 µl of RNA sample at 42 °C for 30 min with 7 units of avian myeloblastosis virus reverse transcriptase (Promega) in 20 µl buffer containing 50 mM-Tris–HCl pH 8·0, 5 mM-$MgCl_2$, 5 mM-DTT, 50 mM-KCl, 0·05 µg/µl BSA, 15% DMSO, 600 µM each of dATP, dCTP, dGTP and TTP, 1·5 µM primer and 10 units RNasin (Promega).

PCR was performed from 1 µl of the cDNA over 25 cycles with each consisting of 25 s at 94 °C, 35 s at 50 °C and 2·5 min at 68 °C. The extension time for the last cycle was increased to 9·5 min. The reactions were carried out with 0·4 units *Taq* polymerase (Northumbria Biologicals) in 20 µl buffer containing 10 mM-Tris–HCl pH 8·8, 50 mM-KCl, 1·5 mM-$MgCl_2$, 0·1% Triton X-100, 33 µM each of dATP, dCTP, dGTP and dTTP and 0·5 µM of each of the outer nested primers. One µl of the reaction mixture was then transferred to a second tube containing the same medium but with the inner pair of nested primers, and a further 25 heat cycles were carried out with the same programme. The PCR products were subjected to electrophoresis in 3% low melting point agarose gel (IBI) and the fragments were detected by ethidium bromide staining and u.v. illumination. For sequence analysis, single molecules of cDNA were obtained at a suitable limiting dilution at which a Poisson distribution of positive and negative results was obtained (Simmonds *et al.*, 1990*a*).

Direct sequencing of PCR products. The PCR products were purified by glass-milk extraction (GeneClean; Bio101). One-quarter of the purified products was used in sequencing reactions with T7 DNA polymerase (Sequenase; United States Biologicals) performed according to the manufacturer's instructions except that the reactions were carried out in 10% DMSO and the template DNA was heat-denatured before primer annealing.

Phylogenetic methods. The sequences were compiled by version 2.0 of the programs of Staden (1984) and analysed by programs available in the University of Wisconsin Genetics Computer Group sequence analysis package, version 7.0 (Devereux *et al.*, 1984). Phylogenetic trees were inferred using two different programs available in the PHYLIP package of Felsenstein (version 3.4, June 1991; Felsenstein, 1988). The program DNAML finds the tree of the highest likelihood (the maximum likelihood tree) given a particular stochastic model of molecular evolution and has been shown to perform well in simulation studies (Saitou & Imanishi, 1989). In the analyses performed here the global (G) option was used as this searches a greater proportion of all possible trees. The second program used was NEIGHBOR which clusters (following the algorithm of Saitou & Nei, 1987) a matrix of nucleotide distances previously estimated using the program DNADIST (which itself was set, using the D option, to use the same stochastic model as underlies DNAML in order to estimate distances corrected for the probabilities of multiple substitution). In all cases the maximum likelihood and neighbour joining procedures produced congruent trees and thus only the former have been presented here.

Table 1. *Sequences and sources of primers used for amplification of HCV genome*

Name	Region	Position of 5′ base*	Polarity†	Sequences 5′ to 3′	Reference
209	5′NCR	8	−	ATACTCGAGGTGCACGGTCTACGAGACCT	Garson *et al.* (1990)
211	5′NCR	−29	−	CACTCTCGAGCACCCTATCAGGCAGT	Garson *et al.* (1990)
939	5′NCR	−297	+	CTGTGAGGAACTACTGTCTT	Okamoto *et al.* (1990*a*)
940	5′NCR	−279	+	TTCACGCAGAAAGCGTCTAG	Okamoto *et al.* (1990*a*)
410	Core	410	−	ATGTACCCCATGAGGTCGGC	
406	Core	−21	+	AGGTCTCGTAGACCGTGCATCATGAGCAC	
288	NS-3	4951	−	CCGGCATGCATGTCATGATGTAT	Simmonds *et al.* (1990*b*)
290	NS-3	4932	−	GTATTTGGTGACTGGGTGCGTC	Simmonds *et al.* (1990*b*)
208	NS-3	4662	+	TCTTGAATTTTGGGAGGGCGTCTT	
207	NS-3	4699	+	CATATAGATGCCCACTTCCTATC	
242	NS-5	8304	−	GGCGGAATTCCTGGTCATAGCCTCCGTGAA	Enomoto *et al.* (1990)
555	NS-5	8227	−	CCACGACTAGATCATCTCCG	
243	NS-5	7904	+	TGGGGATCCCGTATGATACCCGCTGCTTTGA	Enomoto *et al.* (1990)
554	NS-5	7935	+	CTCAACCGTCACTGAACAGGACAT	

* Position of 5′ base relative to HCV genomic sequence in Choo *et al.* (1991).

† Orientation of primer sequence (+, sense; −, antisense).

To establish the interrelationships of the major types of HCV, we have separately analysed several regions of the viral genome that differ in sequence variability and evolutionary constraint. Thus the conclusions drawn from the sequence comparisons are not subject to spurious evolutionary phenomena that may affect a particular region. However, one problem with the analysis presented here was the absence of a viral sequence that was sufficiently distantly related to HCV to serve as an out-group. Thus, although we describe the interrelationships of different sequence variants of HCV, it should be stressed that we have no means of deciding which sequence is ancestral to the others. The trees are thus drawn in the less familiar unrooted form to indicate this. All sequences reported in this publication have been submitted to GenBank (accession numbers D10113 to D10134).

Results

Analysis of the 5'NCR

Samples were obtained from 18 blood donors that were repeatedly reactive in the Abbott second generation EIA and which were confirmed or indeterminate in the Chiron 4-RIBA (E-b1 to E-b18; Follett *et al.*, 1991). HCV sequences present in stored plasma samples from each donor were amplified with primers corresponding to sites in the 5'NCR (Garson *et al.*, 1990; Okamoto *et al.*, 1990*a*) that are well conserved between all known HCV variants (Table 2, sequences 1 to 15, 28). Sequencing of the PCR product, after limiting dilution to isolate single molecules of cDNA before amplification, allowed approximately 190 bp in the centre of the region to be compared with equivalent published sequences (Fig. 1).

Within the sequences, constant as well as variable regions can be found. Six sequences from donors E-b13 to E-b18 closely resembled HCV-1 (sequence no. 1) and the other published sequences 2 to 15 (Table 2), whereas others (E-b9 to E-b12) resembled the recently reported highly divergent K2 and HC-J6 sequences (nos. 24, 27, 28). However, eight sequences (E-b1 to E-b8) appear quite distinct from the others. Division of the sequences into three groups is supported by formal phylogenetic analysis using the maximum likelihood (Fig. 2) and neighbour joining algorithms (data not shown) of the blood donor sequences along with previously published sequences (identified in Table 2). The group labelled 1 contains sequences of HCV with a world-wide distribution (sequences 1 to 15; Table 2), and group 2 contains K2 and J6 sequences (nos. 24, 27, 28). Sequence variability within the three groups is in each case considerably less than that which separates them, and no sequence intermediate between the three groups was found. This tree shows that the third group is equally distinct from group 1 as is group 2. Using the DNAML model, the corrected distances between sequences within each group were in each case less than 3%. Between groups, they ranged from 9% (between groups 1 and 3, and between groups 1 and 2) to 13% (between groups 2 and 3) (Table 3).

Analysis of the NS-5 region

The nucleotide sequence of the NS-5 region has been found to vary significantly between the previously

Table 2. *Source and citation of previously published HCV sequences used in this study*

No.	Type*	Abbreviation	Geographical source	Reference
1	1 (I)	HCV-1	U.S.A.	Choo *et al.* (1991)
2	1 (I)	Pt-1	Japan	Nakao *et al.* (1991) Enomoto *et al.* (1990)
3, 4	1 (I)	H77, H90	U.S.A.	Ogata *et al.* (1991)
5, 6	1 (I)	GM-1, GM-2	Germany	Fuchs *et al.* (1991)
7	1	J1	Japan	Han *et al.* (1991)
8	1	A1	Australia	Han *et al.* (1991)
9	1	S1	S. Africa	Han *et al.* (1991)
10	1	T1	Taiwan	Han *et al.* (1991)
11	1	U18/I24	U.S.A./Italy	Han *et al.* (1991)
12	1 (II)	HCV-J	Japan	Kato *et al.* (1990)
13	1 (II)	HCV-BK	Japan	Takamizawa *et al.* (1991)
14, 15	1 (I, II)	HC-J1, -J4	Japan	Okamoto *et al.* (1990*b*)
16–20	1 (II)	K1, K1-1–K1-4	Japan	Enomoto *et al.* (1990)
21	1 (II)	JH	Japan	Kubo *et al.* (1989)
22	1 (II)	J7	Japan	Takeuchi *et al.* (1990)
23	1 (II)	T3	Taiwan	Chen *et al.* (1991)
24–27	2 (III) 2 (IV)	K2a, K2a-1 K2b, K2b-1	Japan	Nakao *et al.* (1991) Enomoto *et al.* (1990)
28	2 (III)	HC-J6	Japan	Okamoto *et al.* (1991)
29	2	Clone A	Japan	Tsukiyama-Kohara *et al.* (1991)

* Designation of sequences follows the classification described by Enomoto *et al.* (1990). The alternative classification (Houghton *et al.*, 1991) is shown in parentheses (see Discussion).

		−255	−235	−215	−195	−175	−156
3	E-b1	GGCGTTAGTA	CGAGTGTCGT	GCAGCCTCCA	GGACTCCCCC	TCCCGGGAGA	GCCATAGTGG	TCTGCGGAAC	CGGTGAGTAC	ACCGGAATCG	CTGGGGTGAC
3	E-b2	...	.....	.....	....C.....	.....	.....	.....	.....	.....	.....
3	E-b3	...	.....	.....	....C.....	.....	.....	.....	.....	.....	.....
3	E-b4	...	.....	.....	....C.....	.....	.....	.....	.....	.....	.....
3	E-b5	...	.....	.....	....C.....	.....	.....	.....	.....	.....	.....
3	E-b6	...	.....	.....	....C.....	.....	.....	.....	.....	.....	.....
3	E-b7		.....	.....	..C.C.....	.....	.....	.....	.....	.....	.....
3	E-b8		.....	.....	....C.....	.....	.....	.....	.....	.....	.....
2	K2a	.....	T.....A.	.....	..C.C.....	.....	.....	.....	.....	.....T.	.C...AA...
2	HC-J6	.....	T.....	A.....	..C.C.....	.....	.....	.....	.....	.....T.	.C...AA...
2	E-b9		T.....	A.....	..C.C.....	.....	.....	.....	.....	.....T.	.C...AA...
2	E-b10		T.....	A.....	..C.C.....	.....	.....	.....	.....	.....T.	.C...AA...
2	E-b11		T.....	A.....	..C.C.....	.....	.....	.....	.....	.....TA	.C..AAA...
2	E-b12	...	T.....	A.....	..C.C.....	.....	.....	.....	.....	.....T.	.C..AAA...
2	K2b-1	.....	T.....A.	.....	..C.C.....	.....	.....	.....	.....	.....TA	.CG..AA...
1	E-b13	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	E-b14	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	E-b15	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	E-b16		T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	E-b17		T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	E-b18		T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	HCV-1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	Pt-1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	H77	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	H90	.....	T.....	A.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	GM1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	GM2	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	J1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.C...AG...
1	A1	.....	T.....	.....	....C.....	.....A...	.....	.....	.....	.....T.	.C...A....
1	S1	.....	T.....	.....	....C.....	.....A...	.....	.....	.....	.....T.	.CA..A....
1	T1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	U18/I24	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	HCV-J	.....	T.....T..	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	HCV-BK	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	HC-J1	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...
1	HC-J4	.....	T.....	.....	....C.....	.....	.....	.....	.....	.....T.	.CA..AC...

		−155	−135	−115	−95	−75	−62
3	E-b1	CGGGTCCTTT	CTTGGAGCAA	CCCGCTCAAT	ACCCAGAAAT	TTGGGCGTGC	CCCCGCGAGA	TCACTAGCCG	AGTAGTGTTG	GGTCGCGAAA	GGCC
3	E-b2	.....	.....	.....	.....	.....	.....	.....	.....	.....	....
3	E-b3	.....	.....	.....	.....	.....	.....	.....	.....	.....	....
3	E-b4	.....	.....	.....	.....	.....	.....	.....	.....	.....	....
3	E-b5	.....	.....	.....	.....	.....	.....	.....	.....	.....	....
3	E-b6	.....	.....	.....	.....	.....	.....	.....	.....	.....	....
3	E-b7	.....	.....	.....	.....	.....	.....A...	C.....	.....	.....	....
3	E-b8	.....	.....A...	.....	.....	.....	.....	.....	.....	.....	....
2	K2a	T.....	.....TA..	...A...T..	G...G.TC..	.....	.....A..				
2	HC-J6	T.....	.....TA..	...A...T..	G...G.TC..	.....	.....A...	CTG.....	.....C....	...T.....	....
2	E-b9	T.....	.....TA..	...A...T..	G...G.CC..	.....	.....A...	CTG.....	.....C....	...T.....	....
2	E-b10	T.....	.....TA..	...A...T..	G...G.CC..	.....	.....A...	CTG.....	.....C....	...T.....	....
2	E-b11	T.....	.....TA..	...A...T..	GT..G.TC..	.....	.....A...	CTG.....T	.....C....	...T.....	....
2	E-b12	T.....	.....TA..	...A...T..	GT..G.TC..	.....	.....A...	CTG.....T	.....C....	...T.....	....
2	K2b-1	T.....	.....TA..	...A...T..	GT..G.TC..	.....	.....A..				
1	E-b13	.....	.....T...	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	E-b14	.....	.....T...	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	E-b15	.....	.....T...	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	E-b16	.....	.....TT..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	E-b17	.....	.....TT..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	E-b18	.....	.....T...	.....	G..TG..G..	.....	.....	CTG.....	.....	.....	....
1	HCV-1	.....	.....T...	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	Pt-1	.....	.....TA..	.....	G..TG..G..	.....	.....A..				
1	H77	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	H90	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	GM1	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	GM2	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	J1	.....	.....TA..	.....	G..TG..G..	.....C..	.....A..				
1	A1	.....	.....TA..	.....	G..TG..G.C	.....	.....A...	CTG.....	.....	.....	....
1	S1	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	T1	.....	.....TA..	.....	G..TG..G.C	.....	.....A...	CTG.....	.....	.....	....
1	U18/I24	.....	.....TA..	.....	G..TG..G..	.....	.....A...	CTG.....	.....	.....	....
1	HCV-J	.....	.....T...	.....	G..TG..G..	.....	.....	CTG.....	.....	.....	....
1	HCV-BK	.....	.....T...	.....	G..TG..G..	.....	.....	CTG.....	.....	.....	....
1	HC-J1	.....	.....TA..	.....	G..TG..G..	.....C..	.....A...	CTG.....	.....	.....	....
1	HC-J4	.....	.....T...	.....	G..TG..G..	.....	.....	CTG.....	.....	.....	....

Fig. 1. Comparison of nucleotide sequences in the 5'NCR from British blood donors (E-b1 to E-b18) with previously published HCV sequences. Dots indicate identity with sequence of E-b1 (top line); nucleotide substitutions are indicated. Nucleotide numbering corresponds to that described for the prototype HCV-1 sequence (Choo *et al.*, 1991). Source and citation of published sequences are shown in Table 2; phylogenetic groups are indicated in the left column.

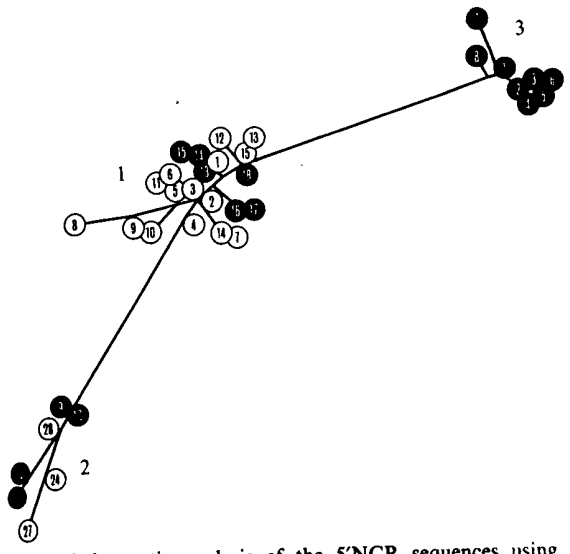


Fig. 2. Phylogenetic analysis of the 5′NCR sequences using the maximum likelihood algorithm, shown as an unrooted tree. ❶ to ⓭ correspond to blood donor sequences E-b1 to E-b18. ① to ㉘ correspond to the previously published sequences identified in Table 2.

described K1 and K2 variants of HCV (Enomoto *et al.*, 1990). To investigate whether the previously identified new group of sequences were equally distant from the other two in this region as well as in the 5′NCR, we compared sequences from four blood donors with group 3 sequences (E-b1, E-b2, E-b3 and E-b7) and one in group 2 (E-b12) with previously published sequences (Fig. 3). As NS-5 is a coding region, the inferred amino acid sequences are presented to indicate the degree of phenotypic variation between the different viral sequences. Phylogenetic analysis of the nucleotide sequences is shown in Fig. 4; corrected nucleotide distances between the different groups are shown in Table 3.

A remarkable variation was observed between sequences of the three groups in this region. Again, sequences falling into the third group clustered separately from the others in this region. However, unlike the 5′NCR, there appear to be subdivisions within the other groups. Group 1 sequences are split between those found in Japanese infected individuals (e.g. HCV-J; HCV-BK; sequence numbers 12, 13, 16 to 20 in Table 2) and those of U.S.A. origin (HCV-1, Pt-1, H77, H90; sequence numbers 1 to 4; Fig. 4). There is also some evidence for a split between group 2 sequences, with K2a and HC-J6 (nos. 24, 25 and 28) appearing distinct from type K2b sequences (nos. 26, 27) and the Scottish blood donor, E-b12.

Table 3 shows that the average nucleotide distances between the two clusters of HCV group 1 sequences is 25% [indicated here as type 1a/I (U.S.A.) and type 1b/II (Japanese)], with variation of only 4 to 7% within each. However, this distance is considerably less than those which exist between group 1 and group 2 sequences (52 to 62%), and group 3 (48 to 49%), and the distance between group 2 and group 3 sequences (53 to 61%).

Table 3. *Nucleotide distances between the three HCV groups in four regions of the genome*

Region	Group*	No.†	1a (I)	1b (II)	2a (III)	2b (IV)	3
5′NCR	1	21	0·0163‡				
	2	7	0·0867		0·0214		
	3	8	0·0948		0·1331		0·0123
Core	1a (I)	6	0·0227				
	1b (II)	5	0·0855	0·0359			
	2a (III)	1	0·2226	0·2051	0·0000		
	3	1	0·1511	0·1802	0·2188	ND§	0·0000
NS-3	1a (I)	34	0·0536				
	1b (II)	4	0·2310	0·0823			
	2a (III)	1	0·3416	0·4072	0·0000		
	2b (?IV)‖	1	0·4337	0·3447	0·2776	0·0000	
	3	4	0·4281	0·3747	0·4885	0·3552	0·0460
NS-5	1a (I)	4	0·0372				
	1b (II)	7	0·2478	0·0743			
	2a (III)	3	0·6170	0·5920	0·0789		
	2b (IV)	3	0·5732	0·5215	0·2328	0·0655	
	3	4	0·4890	0·4755	0·6051	0·5300	0·0322

* See footnote for Table 2.
† Number of sequences analysed.
‡ Groups 1 and 2 could not be subdivided in the 5′NCR, so distances are presented in an intermediate column.
§ ND, Not done.
‖ HCV clone A sequence shown as 2b/IV for presentation purposes only (see text).

		2648	2668	2688	2708	2732
3	E-b1	VTEQDIRVEE	EIYQCCNLEP	EARKVISSLT	ERLYCGGPMF	NSKGAQCGYR	RCRASGVLPT	SFGNTITCYI	KATAACEAAG	LRNPD
3	E-b2									
3	E-b3									
3	E-b7									
2	K2a	...R...T..	S..RA.S.PE	..HIA.H...	....V.....	....QT....	.....T.	.M.....V	..L...K...	IVA.S
2	HC-J6	...R...T..	S..RA.S.PE	..HTA.H...	....V.....	....QT....	.....T.	.M.....V	..L...K...	IIA.T
	K2a-1	...R...T..	S...A.S.PE	...TA.H...	....V.....	....QT....	.....T.	.M.....V	..L...K...	IVA.T
2	E-b12	..	S...A.S.PQ	...T..H...	....V....T	....QS....	.....FT.	.M...M....	..L...K...	IVD.V
	K2b	...R...T..	S...A.S.PQ	...TA.H...	....V....T	....QS....	.....FT.	.M...M....	..L...K...	IVD.I
2	K2b-1	...R...T..	S...A.S.PE	...T..H...	....V....T	....QS....	.....FT.	.M...M....	..L...K...	IVD.I
1	HCV-1	...S...T..	A.....D.D.	Q..VA.K...	....V...LT	..R.EN....	.....T.	.C...L....	..R...R...	.QDCT
1	Pt-1	...S...T..	A.....D.D.	Q..VA.K...	....V...LT	..R.EN....	.....T.	.C...L....	..R...R...	..DCT
1	H77	...S...T..	A.....D.D.	Q..VA.K...	....V...LT	..R.EN....	.....T.	.C...L....	..R...R...	.QDCT
1	H90	...S...T..	A.....D.D.	Q..TA.K...	....V...LT	..R.EN....	.....T.	.C...L....	..R...R...	.QDCT
1	HCV-J	...N...T..	S.....D.A.	...QA.R...	....V...LT	....QN....	.....T.	.C...L...L	.....R..K	.QDCT
1	HCV-BK	...N.....	S.....D.A.	...QA.K...	....I...LT	....QN....	.....T.	.C...L...L	..S...R..K	.QDCT
	K1	...N.....	S...S.D.A.	...QA.R...	....I...LT	....QN....	.....T.	.C...L...L	.....R..K	.QDCT
	K1-1	...N.....	S.....A.A.	...QA.R...	....I...LT	....QN....	.....T.	.C...L...L	..S...R..K	.QDCT
	K1-2	...S.....	S.....D.A.	...Q..R...	....I...LT	....QN....	.....T.	.C...L...L	..S...R..K	..DCT
	K1-3	...N.....	S.....D.A.	..KLA.R...	....I...LT	....QN....	.....T.	.C...L...L	..S...R..K	.QDCT
	K1-4	...N.....	S.....D.A.	...Q..R...	....I...LT	....QN....	.....T.	.C...L...L	..S...R..K	.QDCT

Fig. 3. Comparison of deduced amino acid sequences in the NS-5 region of blood donors E-b1, E-b2, E-b3, E-b7 (type 3) and E-b12 (type 2) with those previously published (Table 2). Amino acid residue numbering follows that of the HCV-1 polyprotein (Choo *et al.*, 1991). Dots indicate identity with sequence of E-b1 (top line); amino acid substitutions are indicated; phylogenetic grouping in the 5′NCR (Fig. 2) is indicated in the left column.

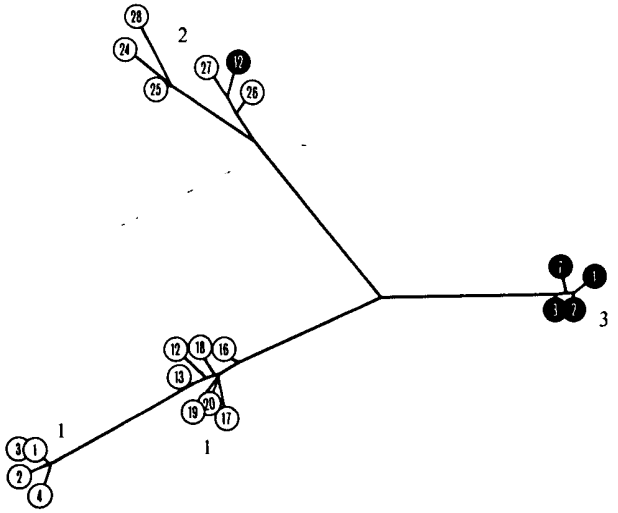


Fig. 4. Phylogenetic analysis of nucleotide sequences in the NS-5 region using the maximum likelihood algorithm, shown as an unrooted tree. Symbols are as described in the legend to Fig. 2.

Analysis of the NS-3 region

Amplification reactions were carried out using previously published primer sequences in the NS-3 region (Weiner *et al.*, 1990), and a pair of empirically derived inner primers (Simmonds *et al.*, 1990*b*). Although these primers amplified HCV sequences from a high proportion of anti-C-100 positive sera from haemophiliacs, they were less effective with sera from IVDUs (Simmonds *et al.*, 1990*b*) and with blood donor samples (three positive out of 15 tested; data not shown). Two conserved sites in the amplified fragment were identified by sequence analysis of the NS-3 region from the haemophiliac and IVDU patients, and two new primers corresponding to these were specified (207, 208; Table 1). The combination of 288 and 208 (first round) and 290 and 207 (second round) primers successfully amplified samples from four donors infected with HCV variants from group 3 (E-b1, E-b2, E-b6 and E-b7) but none of those infected with group 2 sequences (data not shown). This enabled a comparison of sequence group 3 with our own (Simmonds *et al.*, 1990*b*) and previously published sequences (Fig. 5, 6; Table 3). For clarity, only seven of the group 1 sequences obtained in this and our previous study [E-b16, E-b17, i3, i4, h5, h3 and h1 (nos. 19 to 23, filled circles)] are shown in the tree. These sequences are representative of the range of variation found in this region in individuals infected in Britain; comparison of the tree previously published (Simmonds *et al.*, 1990*b*) with Fig. 6 shows that the former forms a very small component of the overall tree obtained once published and group 3 sequences are added.

The maximum likelihood tree again shows three main groups of sequences. As was found in the NS-5 region, sequences in the group 1 5′NCR are split into geographical groups in NS-3. Sequences of Japanese and Taiwanese origin (HCV-J, HCV-BK, JH and T3, nos. 12, 13, 21, 23; Table 2) are distinct from the HCV-1 sequence (no. 1), and those found in Scottish blood donors (E-b16, E-b17, p1 to p3), IVDUs (i1 to i5) and haemophiliacs (h1 to h5) (Fig. 5). However, variation

		1577		1597		1617		1633
3	E-b1	KQQGLNFSYL	TAYQATVCAR	AQALPPSWDE	TWKCLVRLKP	TLHGPTPLLY	RLGPVQN
3	E-b2	....	....	...P......	....	....	....
3	E-b6	....	....	...P......	M.........	....	....
3	E-b7	....	A.........	...P......	....	....	....
2	HC-J6	..S.E..A..	....	.K.P.....V	M....T....	..V.......	...S.T.
	Clone A	..G.D..A..	....	.K.P.....V	M....A....	..I.......	...A.T.
1	E-b16	..S.E.LP..	V.........	...P.....Q	M....I....	....	...A...
1	E-b17	..S.E.LP..	V.........	...P.....Q	M....I....	....	...A...
	Group 1/1	..S.E.LP..	V.........	...P.....Q	M....I....	....	...A...
	1/2	..S.E.LP..	V.........	...P.....Q	M..R.I....	....	.
	1/3	..S.E..P..	V.........	...P.....Q	M....I....	....	..
	1/4	..S.E..P..	V.........	...P.....Q	M....I...A	....	..
	1/4	..S.E..P..	V.........	...P....EQ	M....I...A	....	...A.
1	HCV-1	..S.E.LP..	V.........	...P.....Q	M....I....	....	...A...
1	H77	..S.E..P..	V.........	...P.....Q	M....I....	....	...A...
1	H90	..S.E..P..	V.........	...P.....Q	M....I....	....	...A...
1	HCV-J	..A.D.LP..	V.........	...P.....Q	M....I....	....	...A...
1	HCV-BK	..A.D..P..	V.........	...P.....Q	M....I....	....	...A...
	JH	..A.D..P..	V.........	.K.P.....Q	M....I....	....	...A...
	T3	..A.D..P..	V.........	...P.....Q	M....T....	....	...A...

Fig. 5. Comparison of deduced amino acid sequences in the NS-3 region of blood donors E-b1, E-b2, E-b6, E-b7 (type 3) with those previously published (Table 2). Group 1/1, amino acid sequence of f1, f3, f4, f5, h2, h3, h4 (one), i2, i3, i4, p1, p2; group 1/2, amino acid sequence of i5; group 1/3, amino acid sequence of h2, h3, h4 (one), h5, f2, p3, i1; group 1/4, amino acid sequence of h1 (one); group 1/5, amino acid sequence of h1 (one). Numbering, symbols and abbreviations are as described in the legend to Fig. 3.

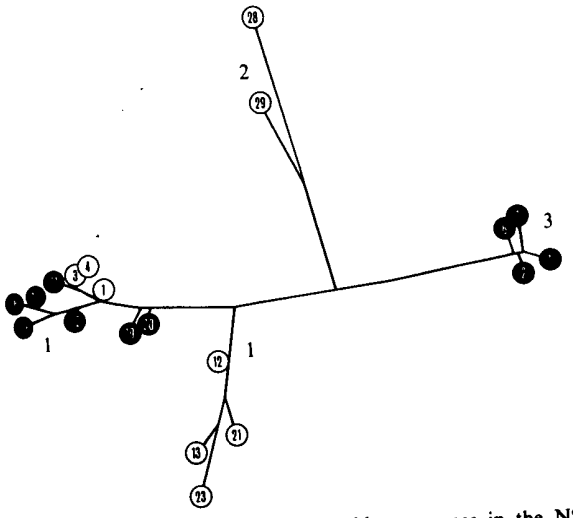


Fig. 6. Phylogenetic analysis of nucleotide sequences in the NS-3 region using the maximum likelihood algorithm, shown as an unrooted tree. Representative nucleotide sequences of the five groups of type 1 sequences shown in Fig. 5 are coded as follows: ● i3; ● i4; ● h5; ● h3; ● h1. Symbols are as described in the legend to Fig. 2.

within group 1 (23%) is lower than those which exist between group 1 and the four group 3 sequences (37 to 43%), and between the group 2 sequence, HC-J6, no. 28 (34 to 41%). As reported previously (Simmonds *et al.*, 1990*b*), the majority of nucleotide substitutions that exist between type 1 sequences are silent (i.e. do not affect the encoded amino acid sequence), but numerous amino acid substitutions exist between sequences in the three main groups (Fig. 5). The analysis of the NS-3 region includes the sequence of clone A, labelled as no. 29 (Tsukiyama-Kohara *et al.*, 1991), which was obtained from Japanese patients with NANB hepatitis, and which was reported to be distinct from existing HCV sequences. In Fig. 6, this sequence appears to cluster with HC-J6 (no. 28; 72% similarity, compared with 57 to 66% and 64% with group 1 and 3 sequences). Although sequence data for clone A are restricted, it is possible that it may be homologous with K2b/E-b12, as the sequence difference between it and HC-J6 is comparable to that between K2b/E-b12 and HC-J6 in the NS-5 region (Table 3).

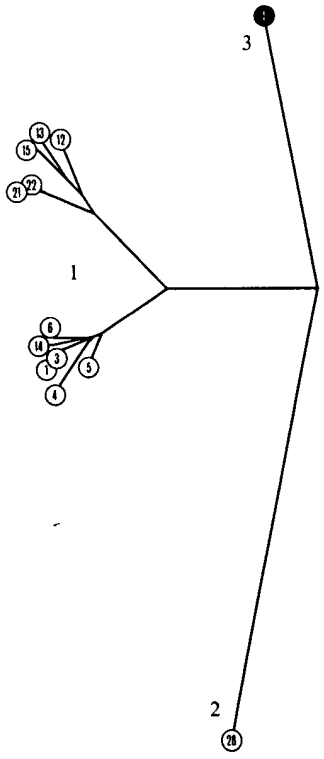
Partial sequence of the putative core region of HCV

The region encoding the putative core protein of HCV is comparatively well conserved in its nucleotide sequence between previously published sequences. Sequences classified in the 5′NCR as group 1 show nucleotide and amino acid sequence similarities of 90 to 98% and 98 to 99% respectively (Fuchs *et al.*, 1991; Ogata *et al.*, 1991), whereas 81% (nucleotide) and 90% (amino acid) simi-

		5 ▼		25 ▼		45 ▼	64 ▼
3	E-b1	AKPQRKTKRN	TIRRPQDVKF	PGGGQIVGGV	YVLPRRGPRL	GVCATRKTSE	RSQPRGRRQP
2	HC-J6	P.....	.N.....	.....	.L.....	..R.....	.....
1	HCV-1	P...K.N...	.N.....	.....	.L.....	..R.....	.....
1	H77	P.....	.N.....	.....	.L.....	..R.....	.....
1	H90	P.....	.N.....	.....	.L.....	..R.....	.....
1	GM1	.....	.N.....	.....	.L.....	..R.P....	.....
1	GM2	.....	.N.....	.....	.L.....	..R.....	.....
1	HCV-J	P.....	.N.....	.....	.L.....	..R.....	.....
1	HCV-BK	P.....	.N.....	.....	.L.....	..R.P....	.....
1	HC-J1	P.....	.N.....	.....	.L.....	..R.....	.....
1	HC-J4	P.....	.N.....	.....	.L.....	..R.....	.....W...
	JH	P.....	.N.....	.....	.L.....	..R.....	.....
	J7	P.....	.N.....	.....	.L.....	..R.....	.....

		65 ▼		85 ▼		105 ▼		128 ▼
3	E-b1	IPKARRSEGR	SWAQPGYPWP	LYGNEGCGWA	GWLLSPRGSR	PSWGPNDPRR	RSRNLGKVID	TLTW
2	HC-J6	...D...T.K	..GK.....	.....L...	.....	.....H	....V.....	...C
1	HCV-1	.....P...	T.....	.....	.....	.....T....	.....	...C
1	H77	.....P...	T.....	.....	.....	.....T....	.....	...C
1	H90	.....P...	T.....	.....M	.....	.....T....	.....	.F.C
1	GM1	.....P...	T.....	.....	.....	.....T....	.....	...C
1	GM2	.....L...	T.....	.....	.....	.....T....	.....	...C
1	HCV-J	.....P...	T.....	.....M...	.....	.....T....	.....	...C
1	HCV-BK	.....P...	T.....	.....L...	.....	.....T....	.....	...C
1	HC-J1	...V..P...	T.....	.....	.....	.....T....	.....	...C
1	HC-J4	.....P...	A.....	.....L...	.....	.....T....	.....	...C
	JH	.....P...	T.....	.....L...	.....	.....	.....	...C
	J7	.....P...	T.....	.....L...	.....	.....	.....	...C

Fig. 7. Comparison of deduced amino acid sequences in the HCV core region of blood donor E-b1 (type 3) with those previously published (Table 2). Numbering, symbols and abbreviations are as described in the legend to Fig. 3.



larities are reported between HC-J6 (group 2) and HCV-1 (group 1; Okamoto *et al.*, 1991). The core region from the blood donor E-b1 (group 3), amplified with primers 410 and 406, was found to be distinct from both group 1 and group 2 (Fig. 7, 8; Table 3). Again there was a prominent subdivision of group 1 sequences into Japanese (HCV-J, HCV-BK, HC-J4, JH and J7; sequences 12, 13, 15, 21, 22) and U.S.A./European (HCV-1, H77, H90, GM1, GM2; nos. 1, 3 to 6, 14) sequences. As was found in NS-3, very little amino acid sequence variation is found between geographically separated variants of group 1 in the core regions; almost all of the nucleotide differences between the two groups are at 'silent' sites. By contrast, HC-J6 shows at least 10 and the E-b1 (group 3) sequence shows at least eight amino acid substitutions in comparison with group 1 sequences.

Fig. 8. Phylogenetic analysis of nucleotide sequences in the core region using the maximum likelihood algorithm, shown as an unrooted tree. Symbols are as described in the legend to Fig. 2.

Discussion

A major difficulty associated with the use of the PCR to assess sequence variation is the possibility that mismatches between the primers and the variant sequence will prevent amplification. In this study, we have used several strategies to overcome this problem. For initial virus detection, we used primers in the 5′NCR, which are reported to be highly conserved amongst previously published sequences (sequences 1 to 15, 28; Table 2). Analysis of the blood donor samples revealed the existence of three phylogenetically distinct groups of sequences. Sequences in the third group have not been previously described. Based on this grouping, we sought corroboration of our findings in other (coding) and more variable regions of the viral genome.

Analysis of the NS-5 region, which was based on several sequences of each of the three main variants (Fig. 3, 4; Table 3), revealed that group 3 sequences formed a relatively homogeneous group that was quite distinct from published sequences. Group 1 and 2 sequences that appeared homogeneous in the 5′NCR split into two distinct clusters. The proposed subdivision of K2 sequences (group 2; Enomoto *et al.*, 1990) into K2a and K2b types is also supported by the phylogenetic analysis presented in this paper. We observed that one of the blood donor sequences obtained in this study appears most similar to K2b, and HC-J6 (Okamoto *et al.*, 1991) falls into the K2a group. Differentiation of HCV type 1 sequences into two groups is also clearly shown in the core (Fig. 7) and NS-3 regions (Fig. 5), in both cases with the group 2 and 3 sequences appearing considerably more distant.

This issue of whether the distinct phylogenetic groupings of HCV constitute separate species cannot be clearly resolved at this stage, as the criteria by which HCV variants may be so differentiated have not yet been defined. Circumstantial evidence in favour of the distinct species hypothesis includes the very marked clustering of phylogenetically distinct groups, without intermediate type sequences. Secondly, we and others have found an overlapping distribution in a single geographical area, and indeed all three groups appear capable of mixed infection in multiply exposed individuals, such as haemophiliacs (Enomoto *et al.*, 1990; Nakao *et al.*, 1991; Chan *et al.*, 1991; Pozzato *et al.*, 1991). Such findings are quite contrary to the alternative hypothesis that the different groups represent epidemiologically separated and hence divergent variants of the same species.

Unlike the 5′NCR, where there are only three distinct groups, each of the coding regions shows prominent differentiation of group 1 sequences into two separate clusters. In this case, the sequences in the two subgroups are geographically separated between Far East (Japan, Taiwan) and U.S.A./European variants (Table 2). However, an exception to this is the HC-J1 sequence (no. 14; Okamoto *et al.*, 1990*b*), which clusters with the U.S.A./European group in the core region (Fig. 8) but was obtained from a Japanese patient with NANB hepatitis. The Pt-1 sequence (no. 2) that also clusters away from the Far East group was obtained from a Japanese haemophiliac treated with imported factor VIII of U.S.A. origin (Enomoto *et al.*, 1990; Nakao *et al.*, 1991). There are insufficient sequence data to indicate whether the two proposed type 2 subtypes, K2a and K2b (Enomoto *et al.*, 1990; Nakao *et al.*, 1991), also represent geographically distinct variants.

There have been several attempts to classify HCV sequence variants into different types. Enomoto *et al.* (1990) classified sequences corresponding to the phylogenetic groups labelled 1 and 2 in this paper as K1/PT and K2, the first being subdivided along the geographical lines discussed previously, and K2 comprising the K2a and K2b subgroups. Houghton *et al.* (1991) classified sequences as type I (sequences of U.S.A./European origin in our phylogenetic group 1), type II (Far East group 1 sequences) and type III (sequences in our phylogenetic group 2), presumably with the option of extending the classification to type IV to differentiate the two clusters of group 2 sequences. However, this classification is uneven in that the difference between type I and II sequences in each protein coding region is consistently less than that between I and III and between I and the new group 3 sequences reported here (Table 3). It is also particularly noticeable that far fewer amino acid substitutions exist between type I and II sequences than between other variants, and type I and II sequences cannot be differentiated in the 5′NCR. For these reasons, we would favour classifying our new sequences in group 3 as type 3, although if the alternative system is accepted, we propose that they be provisionally termed type V.

The genomic organization of HCV corresponds to that of flaviviruses and pestiviruses, with a single open reading frame encoding a polyprotein that is subsequently cleaved into structural and non-structural proteins. Although the overall degree of sequence dissimilarity between the three main groups cannot be measured by comparison of the small regions of sequence analysed in this study, a rough estimate of the extent of divergence in protein coding regions is given by an examination of the divergence of the partial core sequences obtained here. This shows that the difference between HCV group 1, 2 and 3 core regions (approximately 10% amino acid sequence divergence) is comparable to that which exists between different serotypes of the flavivirus, tick-borne encephalitis virus (14%; Mandl

et al., 1988), but lower than that which is found between serotypes of a mosquito-borne flavivirus, dengue fever virus (67%), and the West Nile virus (WNV) subgroup (60%). The 5'NCR sequences of the different members of the WNV subgroup are also considerably more diverse (<50% similarity) than those of the three major types of HCV, although within each of the members, e.g. Murray Valley encephalitis virus, the 5'NCR is extremely well conserved (>95% similarity; Coelen & Mackenzie, 1990). On the basis of these analogies, we speculate that the major types of HCV could conceivably represent distinct serotypes, each capable of human infection irrespective of the immune response mounted against other HCV types.

The existence of different hepatitis C viral types opens up the possibility that the distinct clinical disease syndromes associated with HCV infection may reflect underlying differences in the pathogenicity of the different types of the virus. There is some evidence that infection with HCV group 2 variants (type III) leads to more severe disease than group 1, and is less susceptible to interferon treatment (Pozzato *et al.*, 1991). There are, as yet, no data to link virus type with different sources of infection (particularly non-parenteral infection). Our own preliminary investigations have shown that infection with HCV type 3 is more strongly associated with previous intravenous drug misuse than types 1 and 2 (unpublished data).

The degree of sequence variability found between HCV types would be expected to affect profoundly the antigenicity of many of the putative proteins of HCV. We have previously shown that sera from blood donors infected with different HCV types show distinct differences in the pattern of reactivity to a range of structural and non-structural proteins in two commercial immunoblot assays for HCV antibody (Ortho RIBA and Innogenetics LIA; Chan *et al.*, 1991). In particular, no reactivity with C-100 and infrequent reactivity with C33c were observed in patients infected with HCV types 2 and 3, presumably reflecting the high degree of sequence variability in the NS-3 and NS-4 regions of the genome. Reactivity, however, was always found with the core protein, which is consistent with the degree of sequence conservation in this region (Table 3).

This then provides at least one explanation for the observation that blood donor screening with the original C-100-based immunoassay reduced the incidence but did not entirely prevent post-transfusional NANB hepatitis (Esteban *et al.*, 1990; Japanese Red Cross Non-A, Non-B Hepatitis Research Group, 1991). The use of second generation tests that incorporate core proteins will undoubtedly increase the effectiveness of blood donor screening, although the most effective test for HCV infection would be an assay incorporating representative antigens from all three HCV types. Indeed an immunoblot assay that included polypeptides corresponding to the C-100 protein of HCV types 1, 2 and 3 may serve to type infected individuals serologically, by virtue of the apparent type-specific serological reactivity to this variable protein.

The authors are indebted to staff at the Glasgow and West Scotland Blood Transfusion Service, the Edinburgh and South East Scotland Blood Transfusion Service for supplying blood donor samples, Dr I. Hart of the Scottish Protein Fractionation Centre for plasma pools, Dr G. Kemball-Cook, National Institute for Biological Standards and Control for non-heat-treated commercial factor VIII and Dr C. A. Ludlam, Department of Haematology, Royal Infirmary of Edinburgh for haemophiliac blood samples used for sequence analysis. This work was supported by the Medical Research Council (grant number G9020615CA) and by the Scottish National Blood Transfusion Service.

References

CHAN, S.-W., SIMMONDS, P., MCOMISH, F., YAP, P.-L., MITCHELL, R., DOW, B. & FOLLETT, E. (1991). Serological reactivity of blood donors infected with three different types of hepatitis C virus. *Lancet* **338**, 1391.

CHEN, P. J., LIN, M.-H., TU, S.-J. & CHEN, D.-S. (1991). Isolation of a complementary DNA fragment of hepatitis C virus in Taiwan revealed significant sequence variations compared with other isolates. *Hepatology* **14**, 73–78.

CHOMCZYNSKI, P. & SACCHI, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry* **162**, 156–159.

CHOO, Q. L., KUO, G., WEINER, A. J., OVERBY, L. R., BRADLEY, D. W. & HOUGHTON, M. (1989). Isolation of a cDNA derived from a blood-borne non-A, non-B hepatitis genome. *Science* **244**, 359–362.

CHOO, Q. L., RICHMAN, K. H., HAN, J. H., BERGER, K., LEE, C., DONG, C., GALLEGOS, C., COIT, D., MEDINA SELBY, R., BARR, P. J., WEINER, A. J., BRADLEY, D. W., KUO, G. & HOUGHTON, M. (1991). Genetic organization and diversity of the hepatitis C virus. *Proceedings of the National Academy of Sciences, U.S.A.* **88**, 2451–2455.

COELEN, R. J. & MACKENZIE, J. S. (1990). The 5'-terminal non-coding region of Murray Valley encephalitis virus RNA is highly conserved. *Journal of General Virology* **71**, 241–245.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

ENOMOTO, N., TAKADA, A., NAKAO, T. & DATE, T. (1990). There are two major types of hepatitis C virus in Japan. *Biochemical and Biophysical Research Communications* **170**, 1021–1025.

ESTEBAN, J. I., GONZALEZ, A., HERNANDEZ, J. M., VILADOMIU, L., SANCHEZ, C., LOPEZ TALAVERA, J. C., LUCEA, D., MARTIN VEGA, C., VIDAL, X., ESTEBAN, R. & GUARDIA, J. (1990). Evaluation of antibodies to hepatitis C virus in a study of transfusion-associated hepatitis. *New England Journal of Medicine* **323**, 1107–1112.

FELSENSTEIN, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* **22**, 521–565.

FOLLETT, E. A. C., DOW, B. C., MCOMISH, F., YAP, P. L., HUGHES, W., MITCHELL, R. & SIMMONDS, P. (1991). HCV confirmatory testing of blood donors. *Lancet* **338**, 1024.

FUCHS, K., MOTZ, M., SCHREIER, E., ZACHOVAL, R., DEINHARDT, F. & ROGGENDORF, M. (1991). Characterization of nucleotide sequences from European hepatitis C virus isolates. *Gene* **103**, 163–169.

GARSON, J. A., RING, C., TUKE, P. & TEDDER, R. S. (1990). Enhanced detection by PCR of hepatitis C virus RNA. *Lancet* **336**, 878–879.

HAN, J. H., SHYAMALA, V., RICHMAN, K. H., BRAUER, M. J., IRVINE, B., URDEA, M. S., TEKAMP OLSON, P., KUO, G., CHOO, Q. L. & HOUGHTON, M. (1991). Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the

5′ untranslated region and poly(A) tails at the 3′ end. *Proceedings of the National Academy of Sciences, U.S.A.* **88**, 1711–1715.

HOSEIN, B., FANG, C. T., POPOVSKY, M. A., YE, J., ZHANG, M. & WANG, C. Y. (1991). Improved serodiagnosis of hepatitis C virus infection with synthetic peptide antigen from capsid protein. *Proceedings of the National Academy of Sciences, U.S.A.* **88**, 3647–3651.

HOUGHTON, M., WEINER, A., HAN, J., KUO, G. & CHOO, Q. L. (1991). Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. *Hepatology* **14**, 381–388.

JAPANESE RED CROSS NON-A, NON-B HEPATITIS RESEARCH GROUP (1991). Effect of screening for hepatitis C virus antibody and hepatitis B virus core antibody on the incidence of post-transfusion hepatitis. *Lancet* **338**, 1040–1041.

KATO, N., HIJIKATA, M., OOTSUYAMA, Y., NAKAGAWA, M., OHKOSHI, S., SUGIMURA, T. & SHIMOTOHNO, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proceedings of the National Academy of Sciences, U.S.A.* **87**, 9524–9528.

KOONIN, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197–2206.

KUBO, Y., TAKEUCHI, K., BOONMAR, S., KATAYAMA, T., CHOO, Q. L., KUO, G., WEINER, A. J., BRADLEY, D. W., HOUGHTON, M., SAITO, I. & MIYAMURA, T. (1989). A cDNA fragment of hepatitis C virus isolated from an implicated donor of post-transfusion non-A, non-B hepatitis in Japan. *Nucleic Acids Research* **17**, 10367–10372.

KUO, G., CHOO, Q. L., ALTER, H. J., GITNICK, G. L., REDEKER, A. G., PURCELL, R. H., MIYAMURA, T., DIENSTAG, J. L., ALTER, M. J., STEVENS, C. E., TEGTMEIER, F., BONINO, F., COLUMBO, M., LEE, W.-S., KUO, C., BERGER, K., SCHUSTER, J. R., OVERBY, L. R., BRADLEY, D. W. & HOUGHTON, M. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362–364.

MANDL, C. W., HEINZ, F. X. & KUNZ, C. (1988). Sequence of the structural proteins of tick-borne encephalitis virus (Western subtype) and comparative analysis with other flaviviruses. *Virology* **166**, 197–205.

MILLER, R. H. & PURCELL, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proceedings of the National Academy of Sciences, U.S.A.* **87**, 2057–2061.

MURAISO, K., HIJIKATA, M., OHKOSHI, S., CHO, M. J., KIKUCHI, M., KATO, N. & SHIMOTOHNO, K. (1990). A structural protein of hepatitis C virus expressed in *E. coli* facilitates accurate detection of hepatitis C virus. *Biochemical and Biophysical Research Communications* **172**, 511–516.

NAKAO, T., ENOMOTO, N., TAKADA, N., TAKADA, A. & DATE, T. (1991). Typing of hepatitis C virus genomes by restriction fragment length polymorphism. *Journal of General Virology* **72**, 2105–2112.

OGATA, N., ALTER, H. J., MILLER, R. H. & PURCELL, R. H. (1991). Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proceedings of the National Academy of Sciences, U.S.A.* **88**, 3392–3396.

OKAMOTO, H., OKADA, S., SUGIYAMA, Y., TANAKA, T., SUGAI, Y., AKAHANE, Y., MACHIDA, A., MISHIRO, S., YOSHIZAWA, H., MIYAKAWA, Y. & MAYUMI, M. (1990*a*). Detection of hepatitis C virus RNA by a two-stage polymerase chain reaction with two pairs of primers deduced from the 5′-noncoding region. *Japanese Journal of Experimental Medicine* **60**, 215–222.

OKAMOTO, H., OKADA, S., SUGIYAMA, Y., YOTSUMOTO, S., TANAKA, T., YOSHIZAWA, H., TSUDA, F., MIYAKAWA, Y. & MAYUMI, M. (1990*b*). The 5′-terminal sequence of the hepatitis C virus genome. *Japanese Journal of Experimental Medicine* **60**, 167–177.

OKAMOTO, H., OKADA, S., SUGIYAMA, Y., KURAI, K., IIZUKA, H., MACHIDA, A., MIYAKAWA, Y. & MAYUMI, M. (1991). Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. *Journal of General Virology* **72**, 2697–2704.

POZZATO, G., MORETTI, M., FRANZIN, F., CROCE, L. S., TIRIBELLI, C., MASAYU, T., KANEKO, S., UNOURA, M. & KOBAYASHI, K. (1991). Severity of liver disease with different hepatitis C viral clones. *Lancet* **338**, 509.

SAITOU, N. & IMANISHI, T. (1989). Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biological Evolution* **6**, 514–525.

SAITOU, N. & NEI, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biological Evolution* **4**, 406–425.

SIMMONDS, P., BALFE, P., PEUTHERER, J. F., LUDLAM, C. A., BISHOP, J. O. & BROWN, A. J. (1990*a*). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of Virology* **64**, 864–872.

SIMMONDS, P., ZHANG, L. Q., WATSON, H. G., REBUS, S., FERGUSON, E. D., BALFE, P., LEADBETTER, G. H., YAP, P. L., PEUTHERER, J. F. & LUDLAM, C. A. (1990*b*). Hepatitis C quantification and sequencing in blood products, haemophiliacs, and drug users. *Lancet* **336**, 1469–1472.

STADEN, R. (1984). Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Research* **12**, 521–538.

TAKAMIZAWA, A., MORI, C., FUKE, I., MANABE, S., MURAKAMI, S., FUJITA, J., ONISHI, E., ANDOH, T., YOSHIDA, I. & OKAYAMA, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. *Journal of Virology* **65**, 1105–1113.

TAKEUCHI, K., KUBO, Y., BOONMAR, S., WATANABE, Y., KATAYAMA, T., CHOO, Q. L., KUO, G., HOUGHTON, M., SAITO, I. & MIYAMURA, T. (1990). Nucleotide sequence of core and envelope genes of the hepatitis C virus genome derived directly from human healthy carriers. *Nucleic Acids Research* **18**, 4626.

TSUKIYAMA-KOHARA, K., KOHARA, M., YAMAGUCHI, K., MAKI, N., TOYOSHIMA, A., MIKI, K., TANAKA, S., HATTORI, N. & NOMOTO, A. (1991). A second group of hepatitis C virus. *Virus Genes* **5**, 243–254.

VAN DER POEL, C. L., CUYPERS, H. T., REESINK, H. W., WEINER, A. J., QUAN, S., DI NELLO, R., VAN BOVEN, J. J., WINKEL, I., MULDER FOLKERTS, D., EXEL OEHLERS, P. J., SCHAASBERG, W., LEENTVAAR KUYPERS, A., POLITO, A., HOUGHTON, M. & LELIE, P. N. (1991). Confirmation of hepatitis C virus infection by new four-antigen recombinant immunoblot assay. *Lancet* **337**, 317–319.

WEINER, A. J., KUO, G., BRADLEY, D. W., BONINO, F., SARACCO, G., LEE, C., ROSENBLATT, J., CHOO, Q. L. & HOUGHTON, M. (1990). Detection of hepatitis C viral sequences in non-A, non-B hepatitis. *Lancet* **335**, 1–3.

(*Received 5 December 1991; Accepted 31 January 1992*)

STIC-ILL

NLM
RC845.H45

From: Goldberg, Jeanine
Sent: Friday, December 01, 2000 2:57 PM
To: STIC-ILL
Subject: please pull HCV

1. Proc. Natl. Acad. Sci. U.S.A. 91 (21), 10134-10138 (1994)
2.J. Virol. 66, 3225-3229 (1992)
3. Hepatology 19 (3), 551-553 (1994)
4.J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)

THANK YOU

Jeanine Enewold Goldberg
1655
CM1--12D11
306-5817

Original Articles

Conservation of Hepatitis C Virus 5′ Untranslated Sequences in Hepatocellular Carcinoma and the Surrounding Liver

DEBORAH E. SULLIVAN AND MICHAEL A. GERBER

Department of Pathology and Laboratory Medicine, Tulane University School of Medicine, New Orleans, Louisiana 70112

Persistent infection by hepatitis C virus is a major risk factor for the development of hepatocellular carcinoma, but the mechanism of hepatocarcinogenesis is unknown. To study the association of hepatitis C virus with hepatocellular carcinoma, we sequenced part of the 5′ untranslated region of hepatitis C virus from the tumor tissue and the surrounding nontumorous liver of three patients with hepatocellular carcinoma. No sequence differences between tumor-derived and liver-derived hepatitis C virus isolates were detected. The conservation of the 5′ untranslated region of hepatitis C virus—not only in infected hepatocytes, but also in neoplastic cells—suggests that the regulatory elements at the 5′ terminus of the viral genome play an important role in the pathobiology of hepatitis C virus. (HEPATOLOGY 1994;19:551-553.)

Hepatitis C virus (HCV) is a major pathogenic agent of chronic hepatitis which often leads to cirrhosis and HCC (1). The virus contains a positive-stranded RNA genome of approximately 9,400 nucleotides and presumably replicates through formation of negative-stranded RNA intermediates (2). Several complete sequences of HCV have been reported, revealing a single large translational open reading frame preceded at the 5′ end by a relatively long untranslated region (5′ UTR) (3, 4). This 5′ terminal region is 324 to 340 nucleotides long and represents the most highly conserved sequence among different HCV isolates (5). Recent studies suggest that the 5′ UTR of HCV contains positive and negative translational control elements and may play a role in the pathobiology of chronic HCV infection (6).

Many reports from around the world demonstrated a close association between chronic HCV infection and the development of HCC. The mechanism of malignant transformation, however, is not known. We and others recently demonstrated positive-stranded and negative-stranded HCV RNA sequences, both in chronically infected liver tissue and HCC tissue (7-9). To further elucidate the pathobiology of HCV, we cloned and sequenced part of the 5′ UTR of HCV from the tumor tissue and the surrounding nontumorous liver of three patients with UTR.

PATIENTS AND METHODS

Patients. Three male, HCV-positive patients (62, 52 and 74 yr) with chronic hepatitis C, cirrhosis and HCC were selected for this study (patients 3, 10 and 12 of reference 7). HCV antibody (anti-HCV) was detected in the serum of one patient, and HCV RNA untranslated sequences were present in the liver and tumor tissues of all three patients. In our larger series of patients (7), the correlation between the presence of anti-HCV in serum and HCV sequences in liver was much better ($p < 0.025$) than in the three patients selected for study here. All HBV serological markers were negative. The tissues were obtained from the liver explants at the time of orthotopic liver transplantation.

Extraction of RNA from Liver and Tumor Tissues. Two pathologists selected tissue samples by means of histological examination of frozen sections to represent only liver or only HCC. Then the selected fresh frozen tissues (1 mg) were homogenized in a guanidine solution (4 mol/L guanidium thiocyanate, 25 mol/L sodium citrate [pH 7.0], 0.5% sarcosyl, 0.1 mol/L 2-mercaptoethanol) with a glass homogenizer at 4° C (10). Sodium acetate was added to yield a final concentration of 0.2 mol/L, and the lysate was extracted with phenol/chloroform (1:1). The aqueous phase was collected after centrifugation at 14,000 rpm for 10 min at 4° C and extracted with an equal volume of chloroform. We precipitated RNA from the aqueous phase by adding 1 vol of isopropanol and keeping the sample at −70° C for at least 1 hr. RNA was collected by means of centrifugation at 14,000 rpm for 30 min at 4° C. The RNA pellet was washed with 70% ethanol, air-dried and then resuspended in 10 to 20 μl of sterile, nuclease-free water. RNA samples were aliquoted and stored at −70° C until their use.

Reverse Transcription–Polymerase Chain Reaction. The sequences of the primers for the UTR were as follows: For the first polymerase chain reaction (PCR), the sense primer was 5′-CTGTGAGGAACTACTGTC-3′ and the antisense primer was 5′-CACTACTCGGCTAGCAGT-3′; for the second PCR, the sense primer was 5′-CACGCAGAAAGCGTCTAG-3′ and the antisense primer was 5′-TTGATCCAAGAAAGGACCC-3′. The expected lengths of first- and second-round PCR products were 219 and 142 bases, respectively. The sequence of the oligomer probe used was 5′-AGTATGAGTGTCGTGCAGC-CTCCAGGA-3′.

Received May 4, 1993; accepted September 24, 1993.

This work was supported by National Institutes of Health grant CA54576.

Address reprint requests to: Michael A. Gerber, M.D., Professor and Chairman, Department of Pathology and Laboratory Medicine, Tulane University School of Medicine, 1430 Tulane Avenue, New Orleans, LA 70112.

0270-9139/94 $3.00 + 0 31/1/52132

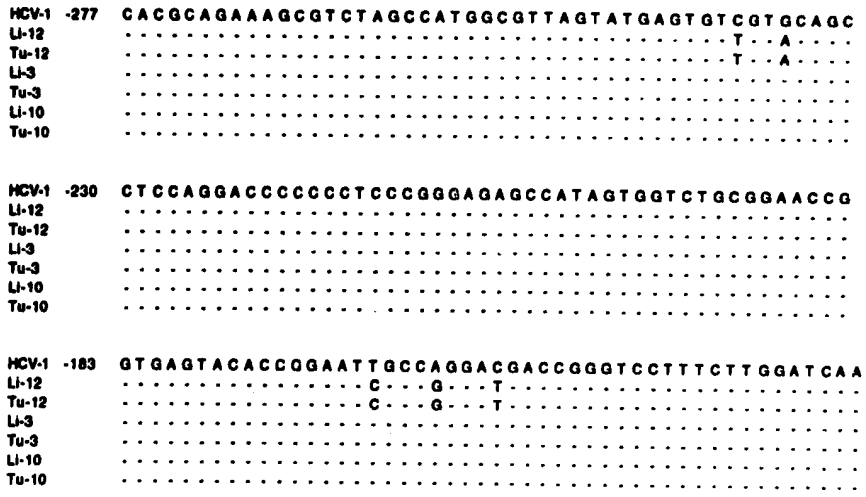


FIG. 1. Nucleotide sequence of the 5′ UTR of the HCV-1 prototype in comparison to sequences of HCV RNA from liver *(Li)* or tumors *(Tu)* of three patients.

We performed first-strand cDNAl synthesis by mixing 500 ng liver or tumor-derived total RNA with 500 ng (80 pmol) of the outer antisense primer, denaturing at 70° C for 10 min and annealing at room temperature (7). Ten units of avian myeloblastosis virus reverse transcriptase (Promega Corp., Madison, WI) was added and incubated at 42° C for 60 min in the presence of 36 units of RNasin (Ambion, Austin, TX), 1 mmol/L each of the four dNTPs (Promega Corp.), and PCR buffer (50 mmol/L KCl, 10 mmol/L Tris-HCl [pH 9.0], 1.5 mmol/L $MgCl_2$, 0.01% gelatin, 0.1% Triton X-100). The cDNA product was boiled for 10 min and chilled on ice, and 10 μl was amplified in a total volume of 50 μl containing 250 ng (40 pmol) of each of the outer primers, 200 μmol/L dNTPs, PCR buffer and 2.5 units of Taq polymerase (Promega Corp.). The reaction was performed in a programmable DNA thermal cycler (Ericomp, San Diego, CA) for 30 cycles consisting of denaturation at 94° C for 1 min, primer annealing at 55° C for 1 min and primer extension at 72° C for 1 min. Five microliters of PCR product from the first amplification was used as template for a second amplification of 30 cycles (same temperature profile as first amplification) with the inner primer pair. After the second amplification, 10 μl of the PCR product was analyzed by means of electrophoresis on a 2% agarose gel containing ethidium bromide and visualized with UV fluorescence. Specificity of the PCR reaction was confirmed with Southern-blot analysis and hybridization to a ^{32}P-labeled HCV-UTR–specific oligomer probe (see above) under stringent conditions.

Cloning of PCR Products. Specific second-round PCR products corresponding to nucleotides −277 to −136 were gel-purified (11) and ligated directly into a T-vector prepared according to the method of Marchuk et al. (12). pBluescript SK II plasmid (Stratagene, La Jolla, CA) was digested with *SmaI* (Promega Corp.) and incubated with Taq polymerase (1 unit/μg plasmid/20 μl volume) in PCR buffer (50 mmol/L KCl, 10 mmol/L Tris-HCl [pH 9.0], 1.5 mmol/L $MgCl_2$, 0.01% gelatin, 0.1% Triton X-100) containing 2 mmol/L dTTP (Promega Corp.) for 2 hr at 70° C. After phenol/chloroform (1:1) extraction and ethanol precipitation, 100 ng of T-vector and 12 ng of PCR product were ligated at 14° C overnight. The ligation mixture was dialyzed against sterile dH_2O for 1 hr in a microdialyzer (13) with two dialysate changes; 1 μl was used to transform competent DH5-alpha *Escherichia coli* by means of electroporation. We identified positive clones by plating on LB, ampicillin, and X-gal plates and screening of white colonies by means of PCR using HCV-UTR–specific primers.

Positive clones were sequenced with Sequenase (U.S. Biochemical Corp., Cleveland, OH) and standard conditions for sequencing of double-stranded plasmid DNA. Sequencing reactions were analyzed on a 6% denaturing sequencing gel.

Direct Sequencing of PCR Products. Second-round PCR products for sequencing were purified by means of gel electrophoresis through low melting point agarose and then eluted (11). The specific band was excised, melted in 3 vol sterile distilled H_2O at 65° C for 10 min and frozen at −70° C. After thawing, the mixture was centrifuged at 14,000 rpm for 10 min to pellet the agarose and the supernatant containing the HCV PCR product was collected.

The double-stranded PCR products were sequenced with Sequenase (U.S. Biochemical Corp.) under the conditions recommended by Casanova et al. (14). Specifically, 0.25 pmol of PCR product was denatured in the presence of 5 pmol PCR primer and reaction buffer (U.S. Biochemical Corp.) by boiling in 10 μl for 10 min. The mixture was immediately cooled at −70° C for 15 sec, rapidly mixed with 5.5 μl labeling mix (U.S. Biochemical Corp.) and incubated at room temperature for 15 sec. We carried out termination reactions by transferring 3.5 μl of the labeling reaction to each of four wells of a microtiter plate containing 2.5 μl of the respective ddNTPs mixture (U.S. Biochemical Corp.) and incubating at 37° C for 2 min. The reaction was stopped with 4 μl of stop solution (U.S. Biochemical Corp.). Sequencing reactions were heat-denatured at 80° C for 3 min and analyzed on a 6% denaturing sequencing gel.

RESULTS AND DISCUSSION

Through a combination of phylogenetic, thermodynamic and biochemical studies, Brown et al. derived a model of the secondary structure of the 5′ UTR of HCV (15). It revealed several secondary structural elements such as hairpins (stem/loop), bubbles and bulges that may bind regulatory proteins (16). In addition, the 5′ UTR sequence represents the most conserved region of

the HCV genome. Therefore we selected a portion of this region for sequence analysis and comparison of the HCV genomes isolated from HCC tissue and the surrounding liver of the same patients. Bukh and coworkers (17) described three domains of significant heterogeneity in the 5′ UTR of HCV. The region sequenced here corresponds to nucleotides −277 to −136 and spans two of these domains. Nevertheless, we emphasize that only 142 nucleotides of the 341 nucleotides of the 5′ UTR were sequenced and that the remainder of the 5′ leader should be examined.

Clearly separable specimens of tumor and liver were available in the three cases selected. The 5′ UTR of the extracted HCV RNA was reverse transcribed and the resulting cDNA was amplified by means of double PCR. All cases contained amplifiable HCV 5′ UTR sequences, as confirmed with Southern-blot analysis using a HCV-specific oligomer hybridization probe. The 142 base pair PCR products derived from the liver and tumor of case 12 were each cloned into pBluescript II SK$^+$ and sequenced. Comparison of the nucleotide sequences demonstrated no differences between tumor- and liver-derived HCV isolates. When compared with that of the HCV-1 prototype (5), the nucleotide sequences were 97% homologous (Fig. 1) and 100% homologous to isolates Z4, Z6, and Z7, as reported by Bukh et al. (16).

PCR products derived from cases 3 and 10 were sequenced directly. Again, we found no difference between tumor- and liver-derived HCV isolates. The nucleotide sequences of HCV isolates from both cases were 100% homologous to the HCV-1 prototype (Fig. 1). Nucleotide sequence information from all HCV isolates was confirmed by means sequencing of the PCR products from three independent reverse transcription–PCR reactions.

These findings suggest that mutations of the 5′ UTR of HCV, which appears to contain major regulatory control elements, do not play a role in malignant transformation of hepatocytes. Furthermore, there is no evidence that the HCV genome is reverse-transcribed and integrated into the host cell genome (7, 8). In contrast, during chronic HBV infection HBV DNA undergoes major alterations and is integrated into the hepatocyte genome during hepatocarcinogenesis (18). Furthermore, the transcriptional transactivator X protein and the truncated pre-S/S gene may play a role in malignant transformation in HBV infection (18).

In a recent study, Paterlini et al. (8) sequenced a hypervariable region of the E2/NS1 gene of HCV from the serum, liver tissue and tumor tissue of two patients. The sequences obtained from each sample were different and demonstrated several mutations in the same patient. The E2/NS1 region exhibits significant diversity among different viral isolates, even those of the same HCV strain. In contrast, our findings indicate that the 5′ UTR is highly conserved among HCV isolates recovered from tumor and liver tissues of the same and different patients. The findings suggest that the regulatory elements of the 5′ UTR may play an important role in the life cycle and persistence of HCV in hepatocytes and during hepatocarcinogenesis.

REFERENCES

1. Houghton M, Weiner A, Han J, Kuo G, Choo Q-L. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. HEPATOLOGY 1991;14: 381-388.
2. Fong T-L, Shindo M, Feinstone SM, Hoofnagle JH, DiBisceglie AM. Detection of replicative intermediates of hepatitis C viral RNA in liver and serum of patients with chronic hepatitis C. J Clin Invest 1991;88:1058-1060.
3. Choo Q-L, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, et al. Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci USA 1991;88:2451-2455.
4. Okamoto H, Kurai K, Okada S-I, Yamamoto K, Lizuka H, Tanaka T, Fukuda S, et al. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 1992;188:331-341.
5. Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, Tekamp-Olson P, et al. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. Proc Natl Acad Sci USA 1991;88:1711-1715.
6. Yoo BJ, Spaete RR, Geballe AP, Selby M, Houghton M, Han JH. 5′ end-dependent translation initiation of hepatitis C viral RNA and the presence of putative positive and negative translational control elements within the 5′ untranslated region. Virology 1992;191:889-899.
7. Gerber MA, Shieh YSC, Shim K-S, Thung SN, Demetris AJ, Schwartz M, Akyol G, Dash S. Detection of replicative hepatitis C virus sequences in hepatocellular carcinoma. Am J Pathol 1992; 141:1271-1277.
8. Paterlini P, Driss F, Nalpas B, Pisi E, Franco D, Berthelot P, Brechot C. Persistence of hepatitis B and hepatitis C viral genomes in primary liver cancers from HBsAg-negative patients: a study of the low-endemic area. HEPATOLOGY 1993;17:20-29.
9. Sherker AH, Twu J-S, Reyes GR, Robinson WS. Presence of viral replicative intermediates in the liver and serum of patients infected with hepatitis C virus. J Med Virol 1993;39:91-96.
10. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal Biochem 1987;162:156-159.
11. Qian L, Wilkinson M. DNA fragment purification: removal of Agarose 10 minutes after electrophoresis. Biotechniques 1991;10: 736-738.
12. Marchuk D, Drumm M, Saulino A, Collins FS. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res 1991;19:1154.
13. Jacobs M, Wnendt S, Stahl U. High-efficiency electrotransformation of *Escherichia coli* with DNA from ligation mixtures. Nucleic Acids Res 1990;18:1653.
14. Casanova J-L, Pannetier C, Jaulin C, Kourilsky P. Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. Nucleic Acids Res 1990;18:4028.
15. Brown EA, Zhang H, Ping L, Lemon M. Secondary structure of the 5′ non-translated regions of hepatitis C virus and pestivirus genomic RNAs. Nucleic Acids Res 1992;20:5041-5045.
16. Nagai K. RNA-protein interactions. Curr Opinion Struct Biol 1992;2:131-137.
17. Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5′ noncoding region of hepatitis C virus. Proc Natl Acad Sci USA 1992;89:4942-4946.
18. Sherman M, Shafritz DA. Hepatocellular carcinoma and the role of hepatitis B virus. In: Cossman J, ed. Molecular genetics in cancer diagnosis. New York: Elsevier, 1990:349-367.

STIC-ILL

NIC
Q11.NB4

From: Goldberg, Jeanine
Sent: Friday, December 01, 2000 2:57 PM
To: STIC-ILL
Subject: please pull HCV

1. Proc. Natl. Acad. Sci. U.S.A. 91 (21), 10134-10138 (1994)
2. J. Virol. 66, 3225-3229 (1992)

3. Hepatology 19 (3), 551-553 (1994)

4. J. Gen. Virol. 73 (Pt 5), 1131-1141 (1992)

THANK YOU

Jeanine Enewold Goldberg
1655
CM1--12D11
306-5817

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10134–10138, October 1994
Biochemistry

Classification of hepatitis C viruses based on phylogenetic analysis of the envelope 1 and nonstructural 5B regions and identification of five additional subtypes

(genotyping/line probe assay/taxonomy)

LIEVEN STUYVER*†, WOUTER VAN ARNHEM*, ANN WYSEUR*, FRANCISCO HERNANDEZ*, ERIC DELAPORTE‡, AND GEERT MAERTENS*

*Innogenetics, Industriepark 7, Box 4, B-9052 Ghent, Belgium; and ‡Institut National de la Santé et de la Recherche Médicale Unité 13, 190 Boulevard MacDonald, 75019 Paris, France

Communicated by Robert H. Purcell, June 1, 1994

ABSTRACT **Genotyping of hepatitis C virus-positive sera by means of a line probe assay indicated that <3% of European samples, but up to 30% of Gabonese sera, could not be classified as either 1a, 1b, 2a, 2b, 3a, 3b, 4c, 5a, or 6a. Such samples were analyzed in the 5′ untranslated region and in the nonstructural 5 (NS5) region. Classification based on phylogenetic analysis of the commonly used 222-bp-long NS5B region was possible for most but not all of the selected sera. Therefore, the core/envelope 1 region (579 bp) and a larger NS5B (340 bp) region were also analyzed. Only the phylogenetic analysis of the 340-bp NS5B region of these newly identified and published isolates provided unambiguous classification into types and subtypes. Furthermore, unequivocal evidence for four subtypes in type 2 and eight subtypes in type 4 was provided. A specific recognition sequence in the 5′ untranslated region was observed for every newly identified subtype. Based on 1830 pair-wise comparisons in NS5B, isolates belonging to the same subtype showed evolutionary distances of <0.127 and isolates of the same type exhibited evolutionary distances of <0.328. These phylogenetic border distances can be conveniently used for classification of hepatitis C virus isolates into types and subtypes.**

Hepatitis C virus (HCV) is thought to be the causative agent of most non-A, non-B hepatitis cases. A very high number of HCV-infected patients develop chronic hepatitis, which often results in liver cirrhosis and occasionally progresses to hepatocellular carcinoma (1). DNA complements of the complete RNA genome of HCV have been cloned (2–5) and show an organization comparable to those of the genomes of pestiviruses and flaviviruses (6). Within the HCV genus, a high degree of sequence heterogeneity exists. Four groups of complete genomes have been reported. HCV-1 (3), HC-J1 (7), and HCV-H (8) belong to a group designated as type I (9), PT (10), or 1a (11, 12). The second group contains at least eight complete genomes: HCV-J (2), HCV-JK1 [GenBank accession no. (Acc.) X61596], HCV-China (Acc. L02836), HCV-T (13), HC-J4 (14), HCV-TA (15), HCV-BK (16), and HCV-JT (15), and was described as either type II (9), K1 (10), or 1b (11, 12). The third group contains only one prototype sequence, HC-J6 (4), described as type III (9), K2a (10), or 2a (11, 12). Group four, called type III (9), type IV (17), type K2b (10), or type 2b (11, 12), is represented by HC-J8 (5). More groups could be identified based on homology calculations in the nonstructural 5 (NS5), the core, and the envelope 1 (E1) region but also based on variations in the 5′ untranslated region (5′ UR). These included type 3a (11, 12) or type IV (9) or type V (17); type 3b or type VI (17); type 4 (12, 18–21); type 5, or type V (9); and type 6 (19–21). A proposal for classification (11, 12) has now been widely accepted by scientists in the HCV variability field (22). We have previously reported a convenient system (12) for detection of the major types and subtypes of HCV based on genotype-specific variations in the 5′ UR. Attempts have already been undertaken to recognize and classify the major genotypes into serotypes (23–26) but further research is needed before serotyping may become widely available, and serological subtyping may be difficult (26).

HCV exists *in vivo* as a mixture of slightly different genomes, called quasispecies (27). The mutation rate was estimated to be 1.92×10^{-3} (H strain) and 1.44×10^{-3} (HC-J4) base substitutions per site per year (14, 28). The 5′ UR and the NS4B region are the most conserved, while the E1 and E2 regions show more variability (14). The evolutionary algorithm underlying phylogenetic analysis is based on the "neutral theory" (29), which claims that mutations occur randomly and are functionally almost neutral. The program DNADIST of the phylogenetic analysis package PHYLIP (30) computes a phylogenetic distance between two different nucleotide sequences, analyzing all positions, including those that apparently did not change. The obtained phylogenetic distance is represented by the sum of horizontal branch lengths from one isolate to another through the phylogenetic tree.

In this study, we selected HCV-seropositive samples from Gabon, Cameroon, Belgium, and The Netherlands, which were subjected to genotype analysis by a line probe assay (LiPA). Aberrantly reactive samples were selected for sequencing and phylogenetic analysis in the 5′ UR, the core/E1 region, and the NS5B region.§ The recognized types and subtypes (including newly identified genotypes 2d and 4e–4h) contained specific sequence motifs in the 5′ UR. The results of the phylogenetic analysis of the core/E1 (C/E1) and NS5B regions were comparable, but only results obtained from the 340-bp NS5B region were completely conclusive.

MATERIALS AND METHODS

Serum Samples. Serum samples from Gabon (GB) and Cameroon (CAM) were collected previously (31), screened for HCV antibodies with Innotest HCV antibody (Ab) II, and confirmed by INNO-LIA HCV Ab II and III (Innogenetics, Antwerp, Belgium). HCV antibody-confirmed samples were subsequently genotyped by LiPA (12). Serum samples from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HCV, hepatitis C virus; 5′ UR, 5′ untranslated region; LiPA, line probe assay; E1, envelope 1; NS, nonstructural; Acc., GenBank accession no.

†To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L29574–L29635).

Belgium were obtained from blood donors (BE90, BE96, BE97, and BE100), which were confirmed anti-HCV positive by INNO-LIA HCV Ab III. Serum BE95 was obtained from a child with leukemia and a history of blood transfusion. Serum samples from The Netherlands (NE91, NE92, and NE93) were obtained from chronic HCV patients.

PCR, LiPA, Cloning, and Sequencing. RNA isolation, cDNA synthesis, PCR, cloning, and LiPA genotyping using biotinylated 5′ UR amplification products were performed as described (12, 24). The core/E1 and NS5B PCR products were used for direct sequencing and cloning. The 5′ UR sequences of NE91, NE92, and NE93 were obtained from a study in which LiPA typing was confirmed by sequencing (32). The sequences of the universal 5′ UR primers and of HCPr54 have been described (12, 24). The sequence of primer HCPr52 (+) was 5′-TTGGGTAAGGTCATCGAT-ACCCT-3′. The core/E1 region was amplified with primers HCPr52-54, and the NS5B region was amplified as described (10).

Phylogenetic Analysis. Previously published sequences were taken from the EMBL/GenBank data base. Corresponding accession numbers are given in the Introduction or in the references in the phylogenetic trees. Alignments were created using the program HCVALIGN (F.H., unpublished data). Sequences were presented in a sequential format to the Phylogeny Inference Package (PHYLIP) version 3.5c (30). Distance matrices were produced by DNADIST using the Kimura two-parameter setting and further analyzed in NEIGHBOR, using the neighbor-joining setting. The program DRAWGRAM was used to create a graphic output.

RESULTS

Serum Selection. Using the LiPA, ≈500 anti-HCV positive serum samples were genotyped. Genotypes 4 and 5 were detected in up to 60% of the Gabonese samples, but these so-called "African genotypes" were also detected in 5–15% of Benelux samples. Based on these LiPA results, <3% of the Benelux, but 12 of 39 (31%) of the Gabonese samples could not be typed or subtyped. However, the presence of HCV RNA was confirmed by means of the LiPA system. The 5′ UR sequence of these exceptional cases was determined; for some selected samples, the core/E1 and NS5B regions were analyzed.

The 5′ UR. Fig. 1 shows the 5′ UR sequences of 16 of the selected isolates compared with prototype sequences. Isolate BE90 did not hybridize with one of the typing probes because of variation at positions −159 and −127. Serum NE92 only reacted with the type 2-specific probes in the LiPA but could not be subtyped on the 2a and 2b probes and was therefore selected for further analysis. BE96 could not be typed because of variations at positions −163 and −122. In total, 10 newly identified type 4-related sequences are shown. Only GB549, GB809, CAM600, CAM736, CAMG22, and CAMG27 could not be recognized by means of the LiPA. GB438 was recognized as type 4 but is different from other type 4 sera due to a variation at position −238. The three other Gabonese (GB48, GB116, and GB358) isolates were included for reasons of comparison.

The E1 and NS5B Regions. Primer set HCPr52-HCPr54 permitted the PCR amplification of DNA fragments from the core/E1 region of 12 sera. Such a PCR fragment includes 64 codons of the core 3′ terminus and 127 codons of E1 with a total length of 579 bp. With the NS5 primer set (10), 340-bp-long PCR fragments were obtained from 14 different sera. PCR fragments were subjected to either cloning, direct sequencing, or both. Fig. 2 shows the alignment of the deduced amino acid sequences of the recombinant core/E1 clones. The previously indicated variable regions V1–V5 and the hydrophobic domain (24) are depicted. The eight cysteine residues and the four N-glycosylation sites are conserved.

Phylogenetic Analysis. Four different groups of sequences were analyzed. Group one contained 71 core/E1 sequences of 579 bp (2485 pair-wise comparisons), group two contained 134 E1 sequences of 384 bp (8911 comparisons), group three contained 61 NS5B sequences of 340 bp (1830 comparisons), and group four contained 115 NS5B sequences of 222 bp (6555 comparisons). The border values of phylogenetic distances of each group are given in Table 1. Only the analysis of the 340-bp NS5B region resulted in three nonoverlapping distance ranges. Phylogenetic trees were created from the DNA distance matrices (Fig. 3 and refs. 33 and 34). Phylogenetic distances obtained from two different regions allowed us to classify isolate NE92 as subtype 2d, serum GB809 and CAM600 as subtype 4e, CAMG22 as subtype 4f, GB549 as subtype 4g, and GB438 as subtype 4h. When comparing direct sequencing results with sequences obtained from recombinant clones, the phylogenetic distance between both was always less than the distance between two different isolates belonging to the same subtype. This was demonstrated for most of the investigated sera in the NS5B region and in the E1 region. Only one exception in serum GB809 was found, where clone GB809_4 was significantly different from the direct sequence and from clone GB809_2. The mean phylogenetic distance of 0.2340 provided evidence that the two core/E1 clones were neither derived from the same quasispecies nor from the same subtype, but rather from different type 4 subtypes; serum GB809 is therefore coin-

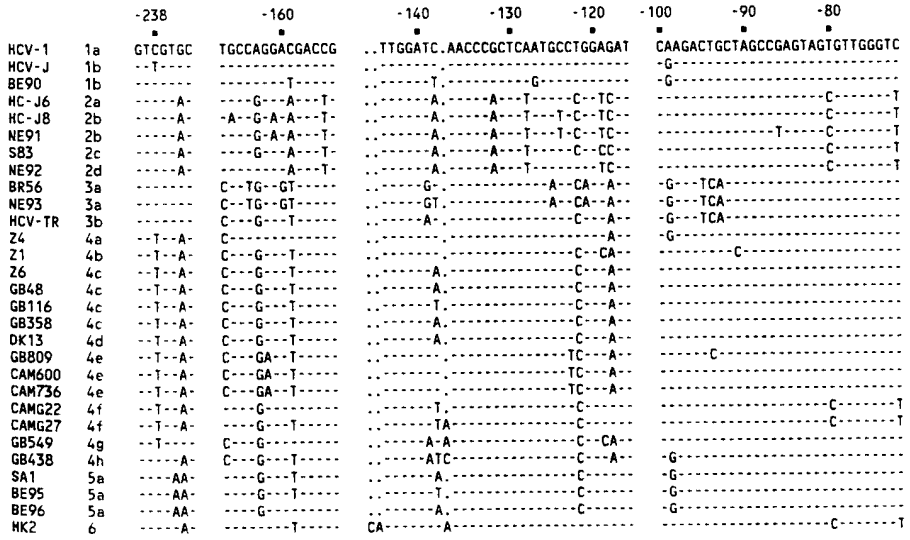


FIG. 1. Nucleotide sequence alignment of the HCV 5′ UR. The sequences between nucleotides −234 and −167, −154 and −144, and −116 and −100 are not shown. Gaps, introduced to preserve alignment, are represented by ". ." and ".". References are given in Fig. 3, except for BR56 (12) and Z1, Z4, Z6, DK13, S83, SA1, and HK2 (20).

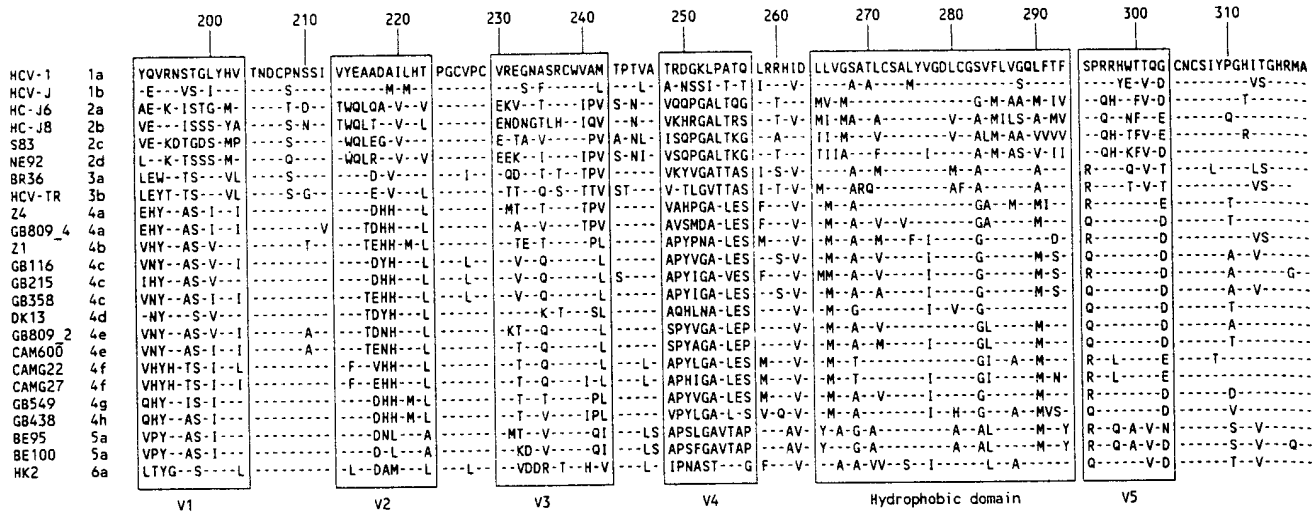


FIG. 2. Deduced amino acid sequence alignment of the HCV E1 region [positions 192–319 (3)]. V1–V5, variable regions 1–5 (24).

fected. Based on the direct sequence of the E1 region, the newly identified subtype 4e was predominant and related to the CAM600 E1 sequences (distance of 0.0637; Fig. 3*A*). As the sequences of the two recombinant NS5 clones (GB809_3_1 and GB809_3_2) are related to the direct sequence of CAM600 (mean distance of 0.0594; Fig. 3*C*), these clones represent subtype 4e of the coinfection.

DISCUSSION

The recently described LiPA technology (12) enabled us to rapidly genotype serum samples obtained from different geographical regions. Sequences of at least 60 5′ UR PCR fragments further supported the reliability of LiPA genotyping (32). HCV genotypes 4 and 5 have hitherto been regarded as the African genotypes. Type 4 has been detected in Zaire (21), in Burundi and Gabon (12), in Egypt (18), and also in Cameroon (this work), whereas genotype 5 was only reported from South Africa (9, 21). Here, we demonstrate considerable prevalences (5–15%) of genotypes 4 and 5 in Europe as well, whereas they could not be found among Brazilian or Japanese sera (ref. 35; L.S., unpublished). Genotype 6 has only been reported from Hong Kong (19, 21).

The use of sequences obtained either from clones or by direct sequencing for subsequent phylogenetic analysis is currently debated. Several studies have employed either method (9, 11, 19, 21), while both sequencing strategies were compared in this study. It could be demonstrated that minor sequence variations, such as those between clones and direct sequences, did not influence the outcome of phylogenetic analysis or classification. However, the potential for misinterpretation from sequencing a single clone from two noncontiguous regions should be acknowledged.

Isolate NE92 was compared in the 5′ UR, the core/E1 region, and NS5 region with type 2a, 2b, and 2c isolates (Fig. 3. This strain was significantly different from type 2a and 2b prototype sequences, but also from type 2c isolate S83 (21) in the E1 region and the type 2c isolates Arg6, Arg8, I10, and T983 (19) in the NS5 region. NE92 can therefore be classified into a newly identified subtype, 2d. Whether type 2c isolates described in E1 (21) and NS5 (19) belong to the same subtype has not yet been demonstrated.

For most subtypes of type 4, specific motifs can again be detected in the 5′ UR. A basic type 4 motif is located between nucleotide positions −240 and −230, and subtype-specific motifs can be observed between positions −170 and −155 and between −125 and −115. From a group of four isolates of subtype 4c, sequences were obtained from the 5′ UR, core/E1 region, and NS5B region, and all sequences clustered together after phylogenetic analysis of each region. Identical subtype 4e motifs (Fig. 1) were found in three different isolates collected from Gabon and Cameroon (GB809, CAM600, and CAM736). According to the phylogenetic analysis of the E1 and NS5B region, they belong to the same subtype (Fig. 3). Similarly, a very specific 5′ UR signature sequence for isolates CAMG22 and CAMG27 in subtype 4f was also found. Subtype-specific motifs were also discovered for subtypes 4g and 4h, but it cannot be predicted whether these signature sequences will be maintained in other isolates belonging to these subtypes.

Because of the large number of complete E1 sequences and the priority of publication, the proposal for type 4 subtype nomenclature of Bukh *et al.* (21) was followed. One of the subtypes of the double-infected serum sample GB809 clustered together with Z4 (subtype 4a) (Fig. 3*A*). Several type 4c isolates (related to Z6 and Z7) were also discovered in this study, but we found no isolates related to Z1 (subtype 4b) or DK13 (subtype 4d). Six other isolates were included into newly identified subtypes 4e (GB809 and CAM600), 4f (CAMG22 and CAMG27), 4g (GB549), and 4h (GB438). A

Table 1. Molecular evolutionary distances

Region	Core/E1, 579 bp	E1, 384 bp	NS5B, 340 bp	NS5B, 222 bp
Isolates	0.0017–0.1347	0.0026–0.2031	0.0003–0.1151	0.0000–0.1323
	(0.0750 ± 0.0245)	(0.0969 ± 0.0289)	(0.0637 ± 0.0229)	(0.0607 ± 0.0205)
Subtypes	0.1330–0.3794	0.1645–0.4869	0.1384–0.2977	0.117–0.3538
	(0.2786 ± 0.0363)	(0.3761 ± 0.0433)	(0.2219 ± 0.0341)	(0.2391 ± 0.0399)
Types	0.3479–0.6306	0.4309–0.9561	0.3581–0.6670	0.3457–0.7471
	(0.4703 ± 0.0525)	(0.6308 ± 0.0928)	(0.4994 ± 0.0495)	(0.5295 ± 0.0627)

Figures created by the PHYLIP program DNADIST are expressed as minimum to maximum (average ± SD). Phylogenetic distances for isolates belonging to the same subtype (Isolates), to different subtypes of the same type (Subtypes), and to different types (Types) are given.

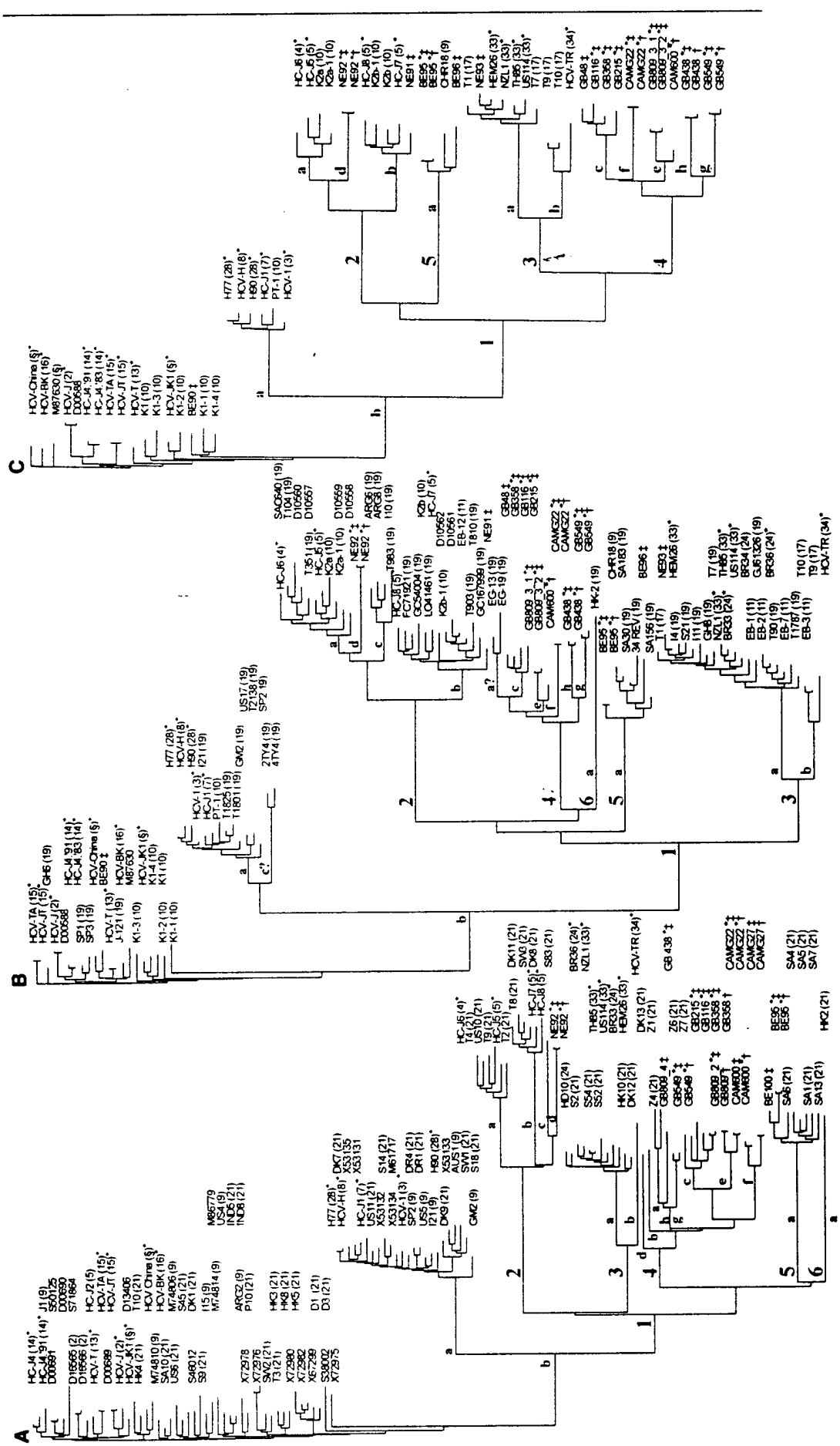


FIG. 3. Phylogenetic tree drawn from the phylogenetic analysis of nucleotide sequences by the neighbor-joining method: 134 sequences of the E1 region (positions 574–957; *A*), 115 sequences of the NS5B region (positions 7975–8196; *B*), and 61 sequences of the NS5B region (positions 7935–8274; *C*). Sequences are indicated with their names and references for accession numbers or directly with their accession numbers. *, Sequence information from the same isolate available in both the core/E1 and NS5B region; §, sequence not published, but accession number given in Introduction; †, sequence obtained by direct sequencing; ‡, one recombinant clone sequenced, corresponds with one accession number.

subtype 4a classification was also assigned to isolates originating from Egypt (18, 19), but it is not clear if EG-29, EG-33, and EG-21 [only partial core sequences published (18)] belong to the same subtype as EG-17, EG-19, and EG-7 [only NS5B sequences published (19)]. Phylogenetic analysis indicated that the EG-19 and EG-7 isolates possibly belong to another type 4 subtype (Fig. 3*B*).

The relationship between isolates, subtypes, and types was expressed by means of the matrix values obtained with the DNADIST program (Table 1), rather than by percentage of homology (see also refs. 11 and 21). This study allowed us to compare the molecular evolutionary distances in two different regions of the HCV genome, as is currently recommended (22). The results obtained in one region generally supported those obtained in other regions. The use of the E1 region is recommended because it provides information with respect to the viral phenotype. However, nonoverlapping evolutionary distances for isolates, subtypes, and types only exist for the 340-bp NS5B region. This finding is contradictory to the previously reported phylogenetic analysis, where the distances of the 222-bp NS5 regions behaved as three nonoverlapping normal distributions (19). Phylogenetic border distances for the 340-bp NS5B region can be used for classification of newly obtained sequences (Table 1). These values may be subject to changes as more sequence data become available. At this moment, intermediate values are set at 0.127 for isolates/subtypes and 0.328 for subtypes/types.

In conclusion, we provide evidence for the existence of at least eight subtypes of type 4 and four subtypes of type 2 and show that, in addition to the E1 region, the phylogenetic analysis should be performed by using nucleotide sequences of at least 340 bp in NS5B.

We thank Ms. M.-C. Dazza (Institut National de la Santé et de la Recherche Médicale Unité 13), Dr. H. Claeys (Bloed Transfusie Centrum, Leuven), Dr. G. Leroux-Roels and Dr. A. Elewaut (State University of Ghent), Dr. R. A. Heijtink and Dr. G. E. M. Kleter (Erasmus University, Rotterdam, The Netherlands), Dr. J. T. Brouwer (University Hospital Dijkzigt, Rotterdam), and Dr. L.-J. van Doorn (Stichting Samenwerkende Delftse Ziekenhuizen, Delft, The Netherlands) for serum collection and providing 5′ UR sequences of NE91, NE92, and NE93. We also thank Dr. J. Louwagie (Innogenetics) and Dr. J. Bukh and Dr. R. Purcell (National Institutes of Health, Bethesda, MD) for helpful discussions and F. Shapiro for manuscript editing. This work was supported by the Flemish Government, Ministry of Economics (Grant FIOV/IWT/91/007.tv1) as part of Eureka Project EU680 "Hepatitis C."

1. Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q.-L., Houghton, M. & Kuo, G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6547–6549.
2. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
3. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
4. Okamoto, H., Okada, S.-I., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. & Mayumi, M. (1991) *J. Gen. Virol.* **72**, 2697–2704.
5. Okamoto, H., Kurai, K., Okada, S.-I., Yamamoto, K., Iizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. & Mishiro, S. (1992) *Virology* **188**, 331–341.
6. Miller, R. H. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
7. Okamoto, H., Kanai, N. & Mishiro, S. (1992) *Nucleic Acids Res.* **20**, 6410.
8. Inchauspe, G., Zebedee, S. L., Lee, D. H. H., Sugitani, M., Nasoff, M. S. & Prince, A. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10292–10296.
9. Cha, T.-A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G. & Urdea, M. S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7144–7148.
10. Enomoto, N., Takada, A., Nakao, T. & Date, T. (1990) *Biochem. Biophys. Res. Commun.* **170**, 1021–1025.
11. Chan, S.-W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. P., Follett, E., Yap, P. L. & Simmonds, P. (1992) *J. Gen. Virol.* **73**, 1131–1141.
12. Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H. & Maertens, G. (1993) *J. Gen. Virol.* **74**, 1093–1102.
13. Chen, P. J., Lin, M. H., Tai, K. F., Liu, P. C., Lin, C. J. & Chen, D. S. (1992) *Virology* **188**, 102–113.
14. Okamoto, H., Kojima, M., Okada, S.-I., Yoshizawa, H., Iizuka, H., Tanaka, T., Muchmore, E. E., Peterson, D. A., Ito, Y. & Mishiro, S. (1992) *Virology* **190**, 894–899.
15. Tanaka, T., Kato, N., Nakagawa, M., Ootsuyama, Y., Cho, M. J., Nakazawa, T., Hijikata, M., Ishimura, Y. & Shimotohno, K. (1992) *Virus Res.* **23**, 39–53.
16. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. & Okayama, H. (1991) *J. Virol.* **65**, 1105–1113.
17. Mori, S., Kato, N., Yagyu, A., Tanaka, T., Ikeda, Y., Petchclai, B., Chiewsilp, P., Kurimura, T. & Shimotohno, K. (1992) *Biochem. Biophys. Res. Commun.* **183**, 334–342.
18. Simmonds, P., McOmish, F., Yap, P. L., Chan, S.-W., Lin, C. K., Dusheiko, G., Saeed, A. A. & Holmes, E. C. (1993) *J. Gen. Virol.* **74**, 661–668.
19. Simmonds, P., Holmes, E. C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993) *J. Gen. Virol.* **74**, 2391–2399.
20. Bukh, J., Purcell, R. H. & Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4942–4946.
21. Bukh, J., Purcell, R. H. & Miller, R. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8234–8238.
22. Simmonds, P., Alberti, A., Alter, H. J., Bonino, F., Bradley, D. W., *et al.* (1994) *Hepatology* **19**, 1321–1324.
23. Machida, A., Ohnuma, H., Tsuda, F., Munekata, E., Tanaka, T., Akahane, Y., Okamoto, H. & Mishiro, S. (1992) *Hepatology* **16**, 886–891.
24. Stuyver, L., Van Arnhem, W., Wyseur, A., DeLeys, R. & Maertens, G. (1993) *Biochem. Biophys. Res. Commun.* **192**, 635–641.
25. Simmonds, P., Rose, K. A., Graham, S., Chan, S.-W., McOmish, F., Dow, B. C., Follett, E. A. C., Yap, P. L. & Marsden, H. (1993) *J. Clin. Microbiol.* **31**, 1493–1503.
26. Maertens, G., Ducatteeuw, A., Stuyver, L., Vandeponseele, P., Venneman, A., Wyseur, A., Bosman, F., Heijtink, R. A. & de Martynoff, G. (1994) in *Viral Hepatitis and Liver Disease*, Proceedings of the 1993 International Symposium on Viral Hepatitis and Liver Disease, eds. Nishioka, K., Suzuki, H., Mishiro, S. & Oda, T. (Springer, Tokyo), pp. 314–316.
27. Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. & Gomez, J. (1992) *J. Virol.* **66**, 3225–3229.
28. Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3392–3396.
29. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
30. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Univ. of Washington, Seattle), Version 3.5c.
31. Delaporte, E., Thiers, V., Dazza, M.-C., Romeo, R., Mlika-Cabanne, N., Aptel, I., Schrijvers, D., Bréchot, C. & Larouzé, B. (1994) *Trans. R. Soc. Trop. Med. Hyg.* **87**, 636–637.
32. van Doorn, L.-J., Kleter, G. E. M., Stuyver, L., Maertens, G., Brouwer, J. T., Schalm, S. W., Heijtink, R. A. & Quint, W. (1994) *J. Hepatol.* **21**, 122–129.
33. Okamoto, H., Tokita, H., Sakamoto, M., Horikita, M., Kojima, M., Iizuka, H. & Mishiro, S. (1993) *J. Gen. Virol.* **74**, 2385–2390.
34. Chayama, K., Tsubota, A., Arase, Y., Saitoh, S., Koida, I., Ikeda, K., Matsumoto, T., Kobayashi, M., Iwasaki, S., Hoyama, S., Morinaga, T. & Kumuda, H. (1993) *J. Gastroenterol. Hepatol.* **8**, 150–156.
35. Stuyver, L., Wyseur, A., Van Arnhem, W., Rossau, R., Delaporte, E., Dazza, M.-C., van Doorn, L.-J., Kleter, G. E. M. & Maertens, G. (1994) in *Viral Hepatitis and Liver Disease*, Proceedings of the 1993 International Symposium on Viral Hepatitis and Liver Disease, eds. Nishioka, K., Suzuki, H., Mishiro, S. & Oda, T. (Springer, Tokyo), pp. 317–319.